Dissertation

Essays on Choice under Uncertainty

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)
eingereicht an der Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Tobias Schmidt

Präsident der Humboldt-Universität zu Berlin:

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Wirtschaftswissenschaftlichen Fakultät:

Prof. Dr. Christian D. Schade

Gutachter:

Prof. Georg Weizsäcker, Ph.D.

PD Dr. Yves Breitmoser

Tag des Kolloqiums:

30.06.2016

Essays on Choice under Uncertainty

Tobias Schmidt

21. April 2016

Acknowledgements

I am deeply indebted to a number of people and institutions for helping to make this dissertation what it is.

First among them is my supervisor and co-author Georg Weizsäcker. Georg not only gave me the freedom to pursue whatever ideas struck my fancy but also helped me in every way imaginable to pursue them. He was there with advice when those ideas developed in ways I had not anticipated or when I ran into problems. He dissuaded me from pursuing some of my weaker ideas but did not hesitate to throw his weight even behind ideas he must have known were risky or unpopular. He was a source of guidance throughout the years, in matters of behavioral economics, the design of laboratory experiments and academic research more generally.

I have also had the great fortune of getting to collaborate with a number of exceptional people on joint research projects, two of which have turned into Chapters 2 and 3 of this dissertation: Steffen Huck, with whom Georg and I collaborated on the series of experiments that make up Chapter 2 and whose wide-ranging interests have been a constant source of inspiration. Charlie Sprenger, whom I met after a seminar talk in Berlin and who invited me to spend two months at Stanford, a time during which parts of Chapter 3, in collaboration with James Andreoni, were written.

Then there is the wider group of people who provided the intellectual ferment from which this thesis draws.

Paul Viefers has been a long-time colleague and close friend. Paul was willing to listen to my ideas at a stage when calling them half-baked would have been complimentary, is one of the most able econometricians I know and a man with a taste and view of the world most similar to mine.

Yves Breitmoser taught me to take the task of confronting theory with data seriously and often challenged me to look at issues through a different lens.

I wrote this dissertation while a member of DIW Berlin's Graduate Center, a graduate program that was in its infancy when I arrived at DIW and which has since developed nicely under the leadership of Georg Meran, Georg Weizsäcker and, most recently, Helmut Lütkepohl. It was the Graduate Center where I met many of the people that have become good friends: Alexandra Avdeenko, Julian Baumann, Lilli Bügelmayer, Ludwig Ensthaler, Nils Saniter, Christoph Grosse Steffen, Jan Marcus, Sören Radde and Lilo Wagner.

For much of my tenure at DIW I was part of the Competition and Consumers department. Special thanks go to the department head Pio Baake.

I owe a debt of gratitude to the Berlin Behavioral Economics community. This group, too, was in its infancy when I arrived in Berlin but has since become one of the largest groups of behavioral and experimental economists in Europe, if not the world. Through its members and the regular series of seminars and workshops I was exposed to cutting-edge research on a staggering variety of topics. I would like to thank Jana Friedrichsen, Paul Heidhues, Tobias Königs and Dorothea Kübler in particular for the many discussions, and for guidance and support over the years.

I thank the European Research Council (Starting Grant 263412) for financial support.

Lastly, I am grateful for my family and my closest friends, Sebastian Königs and Chico Blasques, who have stood by me over the years and, in a very real way, have made me who I am.

Introduction

Mathematical theories of choice under risk and uncertainty are the bedrock on which modern economics is built. Almost all matters that economists concern themselves with involve some element of uncertainty: The value of a good is seldom known at the time of purchase and the consequences of activities as diverse as buying stocks, going to college, choosing an occupation, getting married, having children, smoking or exercising, taking out a life insurance policy or saving for retirement are either not known or unknowable when the decision is made. At the heart of any model of these subjects is a model of how people behave under uncertainty; how they perceive it, what their attitude towards it is and how the two jointly determine decisions.

Von Neumann and Morgenstern (1947) introduced expected utility theory, the first fully formalized theory of choice under risk, in their *Theory of Games and Economic Behavior*.[1] According to expected utility people make decisions as if assigning values — "utilities" — to outcomes, weighting each utility by the objectively given probability with which the outcome will occur and then choosing whichever option gives them the highest weighted, expected utility. Von Neumann and Morgenstern thought of objectively given probabilities that were common to everyone as a crutch and less than a decade later Savage (1954) demonstrated that this crutch was unnecessary. Von Neumann and Morgenstern's theory could be generalized so that every person was free to hold their own, subjective belief and, importantly, that under certain assumptions both people's subjective beliefs and the utilities they attach to different consequences can be inferred from the choices they make.

[1] Though this was the first complete formal treatment of expected utility, precursors can be found in Daniel Bernoulli's analysis of the St. Petersburg paradox in 1738 (Bernoulli, 1954).

Expected utility — in both its objective and subjective version — quickly became the workhorse model of human decision making in modern economics and still holds that status today. But almost immediately after its introduction it came under scrutiny for behavioral assumptions that were obviously at variance with the way people actually behaved. In the seven decades since the publication of *The Theory of Games and Economic Behavior* expected utility has gone through several iterations and restatements and has been generalized in many directions in response to these critiques.

In expected utility the likelihoods with which consequences will materialize are either objectively given or subjective judgments of the decision maker but are then consistently applied to evaluate all options. Some of the generalizations of expected utility break with this consistency requirement. In (cumulative) prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), for example, decision makers may distort the probability of an outcome depending on the rank of the consequence in the gamble and even depending on whether the consequences is seen as a gain or loss relative to a reference point. The salience theory of Bordalo, Gennaioli, and Shleifer (2012) instead posits that some consequences "jump out" at decision makers and therefore carry outsized weight in their decision-making process. Lastly, theories of ambiguity aversion like those of Schmeidler (1989), Gilboa and Schmeidler (1989) or Klibanoff, Marinacci, and Mukerji (2005) posit that there is a fundamental difference between making decisions in situations in which probabilities are known and those in which they are not and that this gives rise to extremely cautious behavior that (subjective) expected utility cannot accommodate. Other theories of choice under risk and uncertainty retain the evaluation of options using a consistent probability measure but instead evaluate consequences differently. In Loomes and Sugden's (1982) regret theory, for examples the consequences of a gamble are evaluated not in isolation but relative to the other consequences that could have materialized: Gaining \$100 may not be quite as happy an occasion if an alternative option would have yielded even more.

While few of the stones from which expected utility is built have been left unturned, almost all mathematical theories of choice under risk and uncertainty have retained one of expected utility's core ideas: In all theories, options are evaluated by a functional that has expected utility form, however unusual either the utility function or the probability measure may sometimes appear to be. People choose among the alternatives by enumerating the possible consequences, making judgments about the likelihoods with which the consequences will occur, assigning to each consequence a subjective value and then assigning to the overall option a value that is a sum of the value of the consequences, each weighted by some function of its likelihood. Common to all theories is therefore a separation of the total value of an option that drives choice into the **utilities** of its constituent parts and decision weights tied to **beliefs** about the likelihood with which those parts will materialize.

The three chapters of this dissertation shine a spotlight on models of choice under uncertainty in this tradition from three different directions. All three chapters are empirical, designed to either study how the theoretical constructs of utility and beliefs can be measured, which of the many theories describes choices best and what explanatory power the theories hold.

Risk Aversion

Chapter 1 takes beliefs as a given and concerns itself with measuring and estimating the curvature of the utility function, which drives aversion to risk in expected utility. The chapter is a re-analysis of Harrison, List, and Towe's (2007) "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," a paper that investigates the sensitivity of experimental procedures for eliciting people's risk preferences to a number of auxiliary assumptions.

In risk preference elicitations it is commonly assumed that subjects will behave the same in the elicitation task — a task that is fashioned to be a close empirical analogue to a theoretical construct, potentially at the cost of artifice — as they would in the real-life situations that are of primary interest to economists. Moreover, it is assumed that both wealth held outside the lab and risks faced outside of the lab ("background risk") play no role in determining the choices made inside the lab. Both assumptions had always seemed questionable on theoretical grounds (see Gollier & Pratt, 1996; Rabin, 2000; Safra & Segal, 2008). Harrison et al. use an artefactual field experiment (Harrisson & List, 2004) to test these assumptions empirically.

The authors recruit numismatists to participate in one of three different elicitation tasks: All three tasks are multiple price list tasks à la Holt and Laury (2002), in which participants face a series of lottery choices. The three tasks differ in the nature of the prizes of the lotteries. The first task is an original Holt and Laury task in which all prizes are monetary. In the second task the monetary prizes are replaced by antique coins in different, known conditions. The authors view these lotteries over coins with which numismatists are intimately familiar as less artificial and abstract than the lotteries over different amounts of money. In the third task, finally, the prizes are the same coins as in the second but with the certificates attesting to the coins' condition removed. In this treatment subjects therefore had to rely on their own expertise to assess the coins' quality and must have felt at least some uncertainty about the true value of the prizes, an additional risk that would remain even after the lottery had played out and which is mathematically equivalent to adding background risk.

Harrison et al. find that varying the artificiality of the experimental task has no statistically significant effect on measured risk preferences nor do subjects behave as if they integrated any of the wealth they hold outside of the lab. Exogenously varying the amount of background risk, however, appears to have an *enormous* influence on measured risk aversion: Being exposed to an additional risk makes subjects much more risk averse.

At the heart of Harrison et al. and at the heart of the re-analysis in Chapter 1 is a measurement issue. The textbook image of choice under risk is a neat one: every person has a utility function, uses this utility function to assign values to prizes and then integrates these values using the given probability measure. He or she does this without fault and whatever option yields the highest expected utility is chosen consistently. Experimental data do not, unfortunately, always conform to this textbook picture. Instead, when asked to make multiple choices, it is often impossible to reconcile all choices with a single utility function. In the analysis of choice data one therefore frequently resorts to stochastic choice models, which allow for non-systematic, random deviations from the "structural" drivers of choice (for an excellent survey of such models, see Wilcox, 2008).[2]

[2] Some have argued that some classic behavioral "anomalies" could be the result of stochastic choice, i.e. while they may appear to be systematic deviations from rationality they may, in

The re-analysis shows that the conclusion that subjects become more risk averse in the face of higher background risk is not robust. It is instead the result of choice data that, for many subjects, could not have been generated by deterministic utility maximization and an econometric model that takes the way in which choices depart from deterministic utility maximization to be evidence of extreme risk aversion. Moreover, the sign, magnitude and statistical significance of the estimated effect of exposing subjects to less artificial prizes is highly dependent on the way one controls for individual heterogeneity in preferences.

Choices in Harrison et al.'s experiment are extremely noisy. One of the choices that subjects in all treatments are presented with, for example, is a "sanity check", a choice between \$200 and \$350, both paid with certainty. For any subject who values money this should be an easy choice. Yet, 47% of the subjects who face this choice choose the lower \$200. Overall, a full 65% of subjects make choices that violate first-order stochastic dominance and are therefore inconsistent with maximizing expected utility deterministically whatever their utility function. In the treatment that identifies the effect of exposing subjects to background risk, choices are particularly noisy. In this treatments subjects appear to be choosing between *any* pair of options put before them entirely at random.

On this data Harrison et al. estimate a "Fechner" stochastic choice model[3] to test whether their experimental treatments have any effect on the amount of risk aversion subjects display. As the Chapter shows, this stochastic choice model is problematic given the extent to which experimental responses depart from deterministic expected utility maximization. Perhaps surprisingly, there are two ways in which the Fechner model can accommodate choices that are completely unsystematic. The first and obvious way is by choosing a choice error so large that any systematic preference is completely drowned out by noise. The second is by

fact, be entirely unsystematic (see e.g. Butler & Loomes, 2007; Collins & James, 2015; Hey & Orme, 1994). Because it creates choices that violate classical postulates like transitivity or monotonicity stochasticity is usually regarded as a mistake. More recently, however, some authors have argued either that there is a higher form of rationality to stochastic choice — choice being stochastic mainly when consequences are of very similar value and when it may not be worth figuring out which option is truly better — or even that there are situations in which people have a strict preference for randomization (see e.g. Agranov & Ortoleva, 2015; Dwenger, Kübler, & Weizsäcker, 2014)

[3] So named because its origins are widely attributed to early work in psychophysics (Fechner, Adler, Howes, & Boring, 1966)

assuming extremely large risk aversion. As is by now well known (Wilcox, 2011), in the Fechner model higher risk aversion is mechanically linked to noisier choices. Asymptotically, both choice errors and high risk aversion imply completely random choice and identical distributions for the data, at which point the model is no longer identified. Chapter 1 shows that depending on the exact specification the data in the background risk treatment are either close to or squarely in this asymptotic case, the reason for the noise being reflected in a high coefficient of relative risk aversion rather than in choice errors being that Harrison et al. do not allow the choice error to differ between treatments.

The extremely noisy data also makes HLT's models numerically unstable and other results highly dependent on model specification. The effect of manipulating the artificiality of the task, negative and not statistically significantly different from zero in the original paper, is statistically significantly positive in other specifications.

Subjective Beliefs

While Chapter 1 takes the utility function as its main object of interest, Chapter 2, based on work with Georg Weizsäcker and Steffen Huck, puts its emphasis on the second component of the expected utility functional: beliefs. The chapter explores the role of subjective beliefs about stock market returns for stock market investment behavior in a representative sample of the German population.

Taking its cue from Savage (1954) most of economics takes data on choices as its empirical primitive. The beliefs that underly these choices, in contrast, are imposed by assumption. Following a seminal paper by Muth (1961) agents are often assumed to hold "rational expectations", beliefs that are objectively correct, at least on average. The reasons for assuming that beliefs are objectively correct are less than compelling and making the assumption regardless may lead to biased tests of other aspects of the theory. As an example, a person assumed to hold rational expectations about stock market returns may appear to be extremely, implausibly risk averse, this being the classical equity premium puzzle result of Mehra and Prescott (1985). But the equity premium would be less of a puzzle if people thought equity returns to be lower or to be more variable than asset pricing models show them to be.[4] As pointed out by Manski (2004), rational expectations are an identifying assumption without which central parameters of interest are not identified. Following Manski a rapidly growing literature has pursued a different strategy to achieve identification: by not only gathering data about people's choices but also asking them for their subjective beliefs.

One way to read Chapter 2 is as probing the properties of the subjective beliefs people report when asked. In particular, the Chapter asks what relationship these reported beliefs bear to the subjective beliefs of SEU. The latter are not beliefs in the common sense meaning of that word but are instead beliefs derived from choice, indeed the beliefs that rationalize choice. Though it may be tempting to take it as a given that the two beliefs are identical, the chapter actually tests this empirically. This test goes further than existing literature, which has found that reported beliefs are predictive of behavior — people who expect the stock market to do better than others tend to hold more stock (see e.g. Dominitz & Manski, 2011; Hurd et al., 2011; Kézdi & Willis, 2009) — but has also uncovered limits to the consistency between beliefs and behavior (see e.g. Merkle & Weber, 2014) and, importantly, has largely neglected issues of causality, i.e. whether people would hold more stock if they believed stock market returns to be more favorable.

The standard portfolio choice problem, a decision task in which a fixed sum of money is to be distributed over two investment vehicles, a risk-free asset that pays a certain rate of return and a risky asset whose return is stochastic, serves as the framework for a series of experiments that explore the determinants of real-world investment behavior. In the main experiment subjects in Germany are given € 25 and confronted with the choice between a German government bond and an asset whose return is tied to historic returns on Germany's blue chip stock market index DAX. An exogenous *shifter* whose sign and magnitude is known to subjects at the time of investment is added to a return randomly chosen from the year-on-year returns on the DAX over the last 60 years.

The experiment then asks subjects to split the € 25 with which they are en-

[4] The empirical evidence on whether beliefs are biased in ways that could explain the equity premium is mixed. Some surveys (Dominitz & Manski, 2011; Hurd, van Rooij, & Winter, 2011; Kézdi & Willis, 2009) find return expectations which are downward biased, others find bias in the opposite direction (Dominitz & Manski, 2011). The survey used in Chapter 2 finds beliefs to be a fairly accurate reflection of the historical distribution of returns.

dowed between these two assets and elicits the beliefs that *should*, under expected utility and more general models, be major determinants of the investment decision. Subjects are asked for their subjective beliefs about the returns on the risky asset they have before them. Using a question format due to Delavande and Rohwedder (2008) and further developed by Delavande, Giné, and McKenzie (2011) we elicit the entire subjective belief distribution.

It's the degree to which beliefs explain investments that is one of the main questions the experiment seeks to answer. Moreover, the study probes the extent to which subjects' beliefs and their investment choices respond to the return shifter. Do subjects offered an investment in a risky asset whose returns are more favorable also expect it to be more favorable? Do they invest more? Lastly, do subjects act similarly in the lab experiment as they do in real life?

The experiment was run with two groups of people: A representative sample of about 700 German households which answered the 2012 survey of the Innovation Sample of the German Socioeconomic Panel, and a group of 198 students at Technical University Berlin.

There are both similarities and differences between these two samples: In both samples return expectations are positively correlated with the amount of money allocated to the risky asset. In the SOEP sample, where information about participants' real-life investments are known, there is also a strong relationship between choices in the portfolio choice experiment and stock market participation: The larger the share of his or her endowment invested in the risky asset, the more likely a participant is to own stocks. Indeed, this "equity share" appears to be an extremely potent predictor of stock holding even after controlling for a number of other determinants that can be found in the literature — gender, age, education, household size, employment status, financial literacy, wealth and income — while, surprisingly, both beliefs and a measure of risk aversion do little to explain participation.

What is particularly interesting, however, are the responses to being treated with different return shifters. In the SOEP sample neither beliefs nor investments respond in any way to the treatment: No matter whether the risky asset pays a past DAX return, a past DAX return minus ten percentage points or a past DAX return plus ten percentage points, on average participants have identical beliefs about returns and allocate the same share of their endowment to the risky asset. Not so in the student sample. Here, too, beliefs do not respond to treatment. Investments, however, do change in the expected direction.

As an explanation for this surprising finding the paper offers a behavioral explanation that departs from subjective expected utility's assumption that options which are payoff-equivalent must be treated identically. In subjective expected utility all that is decision-relevant are the distributions over final outcomes of the available options. The paper shows that while changes to the excess return of an asset should lead to the same change in investments no matter whether the change comes about because the expected return on the safe asset changes or whether the expected return on the risky asset changes, this may not, in fact, be true.

What the Chapter posits instead is that people find it easier to mentally process changes to an object that is relatively simple rather than one that is complex, as suggested by previous lab and field evidence (see e.g Abeler & Jäger, 2015; Chetty, Looney, & Kroft, 2009; Huck & Weizsäcker, 1999; Wilcox, 1993). In a portfolio choice setting, they may find it easier to process changes to an asset that is fully described by a deterministic return rather than one whose return is stochastic. A separate experiment demonstrates that this is indeed the case. In a portfolio choice problem, student subjects respond more strongly to changes to the riskless asset than they do to payoff-equivalent changes in the risky asset.

Beyond Subjective Expected Utility — Ambiguity Aversion

Chapter 3 departs from expected utility and considers more general models of choice under uncertainty. Both the objective expected utility model of von Neumann and Morgenstern (1947) and the subjective expected utility model of Savage (1954) came under attack soon after their formulation. The von Neumann-Morgenstern model was famously challenged by Allais (1953) for assuming that probabilities enter the expected utility functional linearly, an assumption that is violated at certainty and motivates the invention of alternative theories to this day.

The charge leveled against Savage's subjective expected utility was of a different nature: In a series of thought experiments Daniel Ellsberg (1961) demonstrated that people treat situations in which the probabilities of events are not objectively given — genuine uncertainty — differently from situations in which they are — mere risk — in ways that are inconsistent with expected utility. Even Leonard Savage himself showed an aversion to exposing himself to uncertainty that could not be accommodated by a concave utility function. Instead the choice patterns Ellsberg proposed and Savage followed violated one of subjective expected utility's axioms and precluded the existence of a single probability measure that would rationalize all choices. This phenomenon, which has been confirmed in countless laboratory experiments (see Camerer & Weber, 1992, for an overview) has become known as ambiguity aversion and lead to the development of generalizations of subjective expected utility. The earliest such theories — the Choquet expected utility (CEU) model of Schmeidler (1989), the multiple priors or max-min (MEU) model of Gilboa and Schmeidler (1989) and its generalization, α-max-min (Ghirardato, Maccheroni, & Marinacci, 2004) were later followed by the smooth model of Klibanoff et al. (2005, KMM for short) and a number of other models (see Gilboa & Marinacci, 2013, for an excellent survey).

Interestingly, the choice-theoretic literatures on choice under risk and on choice under uncertainty have seen little cross-pollination. The former built on von Neumann and Morgenstern (1947) to accommodate e.g. the Allais paradoxes, either by ascribing to people an anomalous preference for certainty, pessimism or distortions of the objectively given probabilities. The latter departed from Savage to accommodate the Ellsberg paradoxes. The dearth of interaction between these literatures is surprising given that both sets of paradoxes were accommodated by relaxing axioms that bear some similarity.

To understand this, it is useful to look at subjective expected utility through an alternative axiomatic characterization due to Anscombe and Aumann (1963).[5] The objects central to von Neumann and Morgenstern (1947) are "lotteries", objective probability distributions over outcomes. The objects central to Savage are "acts", mappings from states of the world into outcomes. Anscombe and Aumann

[5] Most modern theories on choice under uncertainty are formulated in this Anscombe-Aumann setup.

combine the two into a setup in which payoffs are determined by a two-stage procedure: The first stage is a "horse race", a chance experiment in which outcomes are contingent on the realization of the state of the world for which no probabilities are given. The second stage is a "roulette wheel", a situation in which outcomes are contingent on the outcome of a chance experiment in which — in contrast to the "horse races" above — there are objectively given probabilities. Anscombe-Aumann model decision makers with preferences over mappings from horse races into roulette wheels, which they dub "acts".[6] Every outcome is therefore contingent on an upper level of uncertainty and a lower level of risk. Anscombe and Aumann show that like the Savage axioms imposed on the set of Savage acts, a set of axioms imposed on Anscombe-Aumann acts can be shown to be equivalent to subjective expected utility. Chief among these axioms is "Independence", which states that for any three acts $a, b, c \in F$ and any $\alpha \in (0,1)$

$$a \succsim b \Longleftrightarrow \alpha a + (1-\alpha)c \succsim \alpha b + (1-\alpha)c. \tag{0.1}$$

where F is the set of Anscombe-Aumann acts.

Like vonNeumann-Morgenstern's famous Independence axiom, this axiom imposes on preferences an invariance to probabilistic mixing but does so on a much richer domain. Note that the axiom holds for preferences between *any* pair of acts and for mixing these acts with *any* other act, no matter whether in this act the outcome depends on a "horse race" (i.e. the state of the world), a "roulette wheel" (i.e. the outcome of the objective lottery) or some combination of the two.

All models of ambiguity aversion relax this axiom. However, they do so in different ways. Gilboa and Schmeidler (1989), for example, show that the Ellsberg paradoxes can be accommodated if one relaxes the Independence axiom to an axiom they call Certainty Independence under which preferences are invariant only to mixing with acts that do not depend on the state of the world (these acts are known as constant acts because they yield the same roulette-wheel in every state of the world). Indeed, a broad class of models of ambiguity aversion — the class of invariant biseparable preferences (Ghirardato et al., 2004) which includes

[6] The term "Anscombe-Aumann acts" is often used to distinguish these acts from the "Savage acts" described above, which map states of the world directly into outcomes

(α)-MEU and CEU — all relax Independence in the same way. Other models of ambiguity aversion, most importantly KMM's smooth model, in contrast, relax this axiom further. In KMM, Independence still holds for any triplet of "roulette wheels" but under KMM mixing two "horse races" with a "roulette wheel" does not necessarily preserve preferences.

As shown by Epstein (2010) and Epstein and Schneider (2010) the theories differ fundamentally in the way they conceptualize the "hedging"-value of constant acts. In the class of invariant biseparable preferences mixing an acts with its certainty equivalent is not valued by the decision maker (this is a direct consequence of Certainty Independence, the certainty equivalent being a constant act) because these mixtures do not insure the decision maker against the state of the world but only reduce his exposure to it. KMM, in contrast, conceptualizes a decision maker as being averse to dispersion in state expected utilities. Since mixing with an act's certainty equivalent reduces this dispersion the decision maker values it over and above the reduction in exposure. Epstein and Schneider shows that Certainty Independence is central to this difference: Imposing Certainty Independence on top of KMM yields subjective expected utility.

Probabilistic mixtures of "horse races" with "roulette wheels" are therefore a domain on which theories of ambiguity aversion make differential predictions. Chapter 3, based on work with James Andreoni and Charles Sprenger, provides a *direct* experimental test of the Certainty Independence axiom by studying the valuations that experimental subjects have for particular probabilistic mixtures.

In the experiment subjects are confronted with a source of uncertainty, an 2-color Ellsberg urn. They are then presented with a series of urns designed to be empirical analogues to probabilistic mixtures involving the Ellsberg urn: Each urn contains a combination of balls that lead to an immediate monetary payoff and balls that trigger a draw from the Ellsberg urn to determine the payoff. Varying the share of balls that trigger a draw from the Ellsberg urn varies subjects' exposure to uncertainty. For the particular kinds of valuations the experiment elicits — lottery equivalents as first introduced by Roth and Malouf (1979) — the class of invariant biseparable preference theories predict that valuations be a linear function of the proportion balls in the urn that trigger a draw from the Ellsberg urn. For KMM, valuations ought to be non-linear functions but functional forms that are popular in applications of the theory also imply particular kinds of non-linearity.

The study finds average valuations that are inconsistent with invariant biseparable preferences. Valuations are also at variance with the predictions KMM makes under common assumptions on functional form. Since none of the most popular theories of ambiguity aversion can rationalize the experimental data, the chapter closes with suggestions for properties that might prove useful in coming up with a satisfactory theoretical framework that can.

References

Abeler, J. & Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, *7*(3), 1–28.

Agranov, M. & Ortoleva, P. (2015). *Stochastic choice and preferences for randomization.*

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole americaine. *Econometrica*, *21*(4), 503–546.

Anscombe, F. & Aumann, R. (1963). A definition of subjective probability. *The Annals of Mathematical Statistics*, *34*(1), 199–205.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, *22*(1), 23–36.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, *127*(3), 1243–1285.

Butler, D. J. & Loomes, G. C. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, *97*(1), 277–297.

Camerer, C. & Weber, M. (1992). Recent developments in modeling preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: theory and evidence. *American Economic Review*, *99*(4), 1145–1177.

Collins, S. M. & James, D. (2015). Response mode and stochastic choice together explain preference reversals. *Quantitative Economics*, *6*(3), 825–856.

Delavande, A., Giné, X., & McKenzie, D. (2011). Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? *Journal of Applied Econometrics*, *497*(3), 479–497.

Delavande, A. & Rohwedder, S. (2008). Eliciting subjective probabilities in internet surveys. *Public Opinion Quarterly*, *72*(5), 866–891.

Dominitz, J. & Manski, C. F. (2011). Measuring and interpreting expectations of equity returns. *Journal of Applied Econometrics*, *26*, 352–370.

Dwenger, N., Kübler, D., & Weizsäcker, G. (2014). Flipping a coin: theory and evidence. *CESifo Working Papers*, *4740*.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, *75*(4), 643.

Epstein, L. G. (2010). A paradox for the "smooth ambiguity" model of preference. *Econometrica*, *78*(6), 2085–2099.

Epstein, L. G. & Schneider, M. (2010). Ambiguity and asset markets. *Annual Review of Financial Economics*, *2*(1), 315–346.

Fechner, G. T., Adler, H. E., Howes, D. H., & Boring, E. G. (1966). *Elements of psychophysics*. A Henry Holt Edition in Psychology. Holt, Rinehart and Winston.

Ghirardato, P., Maccheroni, F., & Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, *118*(2), 133–173.

Gilboa, I. & Marinacci, M. (2013). Ambiguity and the bayesian paradigm. In D. Acemoglu, M. Arellano, & E. Dekel (Eds.), *Advances in economics and econometrics: theory and applications, tenth world congress of the econometric society* (Vol. 1, pp. 179–242). New York: Cambridge University Press.

Gilboa, I. & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*, 141–153.

Gollier, C. & Pratt, J. (1996). Risk vulnerability and the tempering effect of background risk. *Econometrica*, *64*(5), 1109–1123.

Harrison, G. W., List, J. A., & Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: a case study of risk aversion. *Econometrica*, *75*(2), 433–458.

Harrisson, G. W. & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(December), 1009–1055.

Hey, J. & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*(6), 1291–1326.

Holt, C. & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Huck, S. & Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: results of a lottery choice experiment. *Journal of Economic Psychology*, *20*, 699–715.

Hurd, M. D., van Rooij, M., & Winter, J. (2011). Stock market expectations of dutch households. *Journal of Applied Econometrics*, *26*, 416–436.

Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Kézdi, G. & Willis, R. J. (2009). *Stock market expectations and portfolio choice of american households*.

Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, *73*(6), 1849–1892.

Loomes, G. & Sugden, R. (1982). Regret theory an alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805–824.

Manski, C. C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329–1376.

Mehra, R. & Prescott, E. C. (1985). The equity premium: a puzzle. *Journal of Monetary Economics*, *15*(2), 145–161.

Merkle, C. & Weber, M. (2014). Do investors put their money where their mouth is? stock market expectations and investing behavior. *Journal of Banking & Finance*, *46*, 372–386.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*(3), 315–335.

Rabin, M. (2000). Risk aversion and expected-utility theory: a calibration theorem. *Econometrica*, *68*(5), 1281–1292.

Roth, A. E. & Malouf, M. W. (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review*, *86*(6), 574–594.

Safra, Z. & Segal, U. (2008). Calibration results for non-expected utility theories. *Econometrica*, *76*(5), 1143–1166.

Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley and Sons.

Schmeidler, D. (1989). Probability and expected utility without additivity. *Econometrica*, *57*(3), 571–587.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

von Neumann, J. & Morgenstern, O. (1947). *The theory of games and economic behavior*. Princeton: Princeton University Press.

Wilcox, N. T. (1993). Lottery choice: incentives, complexity and decision time. *Economic Journal*, *103*(421), 1397–1417.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison. In J. C. Cox & G. W. Harrison (Eds.), *Risk aversion in experiments, research in experimental economics* (Vol. 12, pp. 197–292). Research in Experimental Economics. Emerald Group Publishing Limited.

Wilcox, N. T. (2011). 'stochastically more risk averse:' a contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, *162*(1), 89–104.

Contents

1 Comment on "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion"

The experimental elicitation of risk preferences has received an increasing amount of attention over the past 20 years. The literature has explored many issues, from experimental procedures and stakes (Andersen, Harrison, Lau, & Rutström, 2006; Holt & Laury, 2002) to the econometric methods with which the choice data from such experiments are analyzed (Hey & Orme, 1994; Wilcox, 2008). Harrison, List, and Towe (2007, HLT henceforth) contribute to this literature an artefactual field experiment that investigates the sensitivity of elicited risk preferences to the presence of background risk and to the artificiality of standard elicitation tasks.

Background risk is usually unobserved by the experimenter and econometrician but may influence risk taking in the laboratory in an unknown direction. Expected utility maximizers with decreasing absolute risk aversion, for example, demand higher risk premia after being endowed with an unfavorable independent risk ("risk vulnerability", see Gollier & Pratt, 1996). Not controlling for this risk leads to overestimation of risk aversion. Some psychological theories like diminishing sensitivity, in contrast, predict the opposite effect. Exogenously varying subjects' background risk affords HLT an opportunity to determine both the direction and size of this potential bias. They find that adding background risk makes subjects much more risk averse.

Whether an experiment, in which tasks are often highly stylized, manages to capture the way subjects behave in their daily lives, in the situations that of primary interest to economists is the second question HLT to which devote themselves. By variously using objects with which subjects are more or less familiar

HLT investigate whether the artificiality of standard laboratory procedures has an influence on measured risk preferences and answer the question in the negative.

This comment questions HLT's statistical inference and interpretation. A more detailed look at the experimental data and econometric model shows that the conclusion that exposure to background risk increases risk aversion rests on an experimental treatment in which experimental responses are indistinguishable from having been made entirely at random and an econometric model that identifies this noise as extreme risk aversion. Though the estimated model may seem to be "expected utility plus noise" HLT's parameter estimates imply that the structural component of the model — expected utility — plays no role in fitting a large share of the data. The conclusions that the artificiality of the elicitation task does not measurably influence risk aversion, though based on more informative data, depends on how individual heterogeneity in preferences is modeled.

1.1 Summary

HLT elicit the risk preferences of 113 numismatists at a coin convention using a multiple price list design à la Holt and Laury (2002). In a multiple price list subjects are asked to consider a series of binary choices between a "safe" binary lottery A and a comparatively "risky" binary lottery B, whose prizes are a spread over those in A. In the first row of the list both lotteries pay their low prize with very high probability and their high prize with the small complementary probability. Since lottery A's low prize is higher than B's only the most risk-loving subjects choose B. As one moves down the list the probability of receiving the high prize — always identical for both lotteries — increases steadily. In the last row subjects face the choice between the two high prizes with certainty, a choice in which every expected utility maximizer, no matter his risk attitude, should choose lottery B.[1]

What makes this method appealing and has undoubtedly contributed to its widespread use is that under certain assumptions it allows for a direct mapping from experimental responses to preference parameters. For any (deterministic) ex-

[1] In fact, this holds not just for expected utility but for all preferences that satisfy monotonicity.

pected utility maximizer with strictly increasing utility function there is a unique row in the list at which the subject switches from the safe to the risky lottery. Under additional assumptions on functional form, say that subjects have constant relative risk aversion, the observed switching point identifies the CRRA parameter up to an interval. Tables thus associating switching points with preference parameters abound in the literature.

HLT use three variations of this design. In the *money* treatment all lottery prizes are monetary, the safe lottery A yielding either \$125 or \$200 and the risky lottery B yielding either \$40 or \$350. In the *graded coins* treatment the lottery prizes are 1879-S Morgan silver dollars of different but known quality grades whose retail value was identical to the monetary prizes in the *money* treatment. The *ungraded coins* treatment, finally, features the same prizes as the *graded coins* treatment but with the grading information withheld, the aim being to add uncertainty to the value of each prize. Mathematically this is equivalent to endowing the subject with background risk.

Instead of directly inferring subjects' risk preference parameters from experimental responses HLT use more sophisticated methods (Harrison & Rutström, 2008; Hey & Orme, 1994). They estimate a stochastic choice model according to which each pair of gambles in the list is evaluated using a CRRA utility function but choice is perturbed by i.i.d. normal ("Fechner") errors:

$$\Pr(\text{choice} = A) = \Phi\left(\frac{E\left(\frac{m_A^{1-r_i}}{1-r_i}\right) - E\left(\frac{m_B^{1-r_i}}{1-r_i}\right)}{\sigma_i}\right) \tag{1.1}$$

where $m_j, j = A, B$ are the prizes of lotteries A and B, r_i is the CRRA coefficient of subject i, $E\left(\frac{m_j^{1-r_i}}{1-r_i}\right)$ are the expected utilities, σ_i is the standard deviation of the choice error and $\Phi(\cdot)$ is the CDF of the standard normal distribution. r_i is allowed to vary between subjects according to treatment and various individual characteristics. The variance of the error differs only by subjects' gender.

HLT estimate substantial risk aversion in the *money* and *graded coins* treatments and a difference between the two that is not statistically different from zero (See

Parameter	Variable	Estimate	Std. Error	p-Value	95% Confidence Interval	
r	Constant	0.951	0.444	0.032	0.080	1.822
	Coins As Final Outcomes	-0.160	0.610	0.793	-1.355	1.036
	Ungraded Coins As Final Outcomes	3.974	0.744	0.000	2.517	5.431
	Frame to skew RA lower	0.756	0.399	0.058	-0.026	1.538
	Frame to skew RA higher	0.142	0.299	0.634	-0.443	0.728
	Female	-1.259	0.711	0.077	-2.653	0.135
	College education or higher	0.044	0.209	0.832	-0.366	0.455
	Single and never married	0.838	0.379	0.027	0.094	1.581
	Ever owned Morgan Silver dollars	0.032	0.573	0.956	-1.091	1.155
	Coin dealer	-0.984	0.596	0.099	-2.151	0.184
	Dealer X coins	0.394	0.472	0.405	-0.532	1.319
	Affiliated with a grading company	-0.124	0.283	0.661	-0.680	0.431
σ	Constant	0.408	0.734	0.578	-1.030	1.847
	Female	12.537	38.277	0.743	-62.485	87.559

Table 1.1: Results for the main model (reproduced from Table II in HLT)

Table 1.1). Using monetary instead of "natural" prizes, in other words, has no discernible effect on estimated risk preferences. Presenting subjects with *ungraded coins* instead of *graded coins*, however, has a sizable effect on the estimated CRRA parameter. The estimate on the *ungraded coins* treatment dummy is 3.974 and highly statistically significant. While subjects in the other two treatments are estimated to have average CRRA parameters of 0.88 and 0.77 respectively, the average in the *ungraded coins* treatment is 4.78. HLT interpret this as a drastic increase in risk aversion and as strong evidence of risk-vulnerability.

1.2 Reanalysis

A reanalysis of the data, however, reveals several problems with the original analysis, some related to its econometric particulars, others of a more conceptual nature.

1.2.1 Signal vs. Noise

First, relative to the scale of the utility function assumed throughout the paper[2] the estimated standard deviations of the choice error, $\sigma_{Constant}$ and σ_{Female}, are 100 times larger than those reported in Table II of the paper. This is because in the specification that HLT estimate and whose parameters they report the utility

[2] The scale implied by HLT's equations (1)–(5)

function is scaled down by a factor of 100.[3] This reduces the estimated error standard deviations one to one. While the scaling of the utility function is arbitrary, the size of the choice error relative to whatever scaling is chosen is an important metric for evaluating the fit of the model. And while the stochastic component of the model seems to play only a minor role under the reported parameters its role is, in fact, major. For more than 3/4 of all subjects *nothing but* the stochastic component of the model plays any role in fitting the data.

Second, the coefficient for the treatment indicator of the *ungraded coins* treatment does not maximize the likelihood function and, best as can be told given the limits of numerical precision, the likelihood may not have a maximum at all. Figure 1.1 shows slices of the log-likelihood function for all 14 parameters and indicates that while none of the other parameters can be changed to improve the log-likelihood at the reported maximum, the log-likelihood has a positive albeit small slope at the reported *ungraded coins* parameter.

What produces the estimate of 3.974 given this relatively flat likelihood is the tolerance level at which the Newton-Raphson maximization algorithm declares convergence. Lowering the tolerance on the scaled gradient (the main convergence criterion in Stata 12) from its default of 10^{-5} to 10^{-8} changes the parameter to 316.638.[4][5]

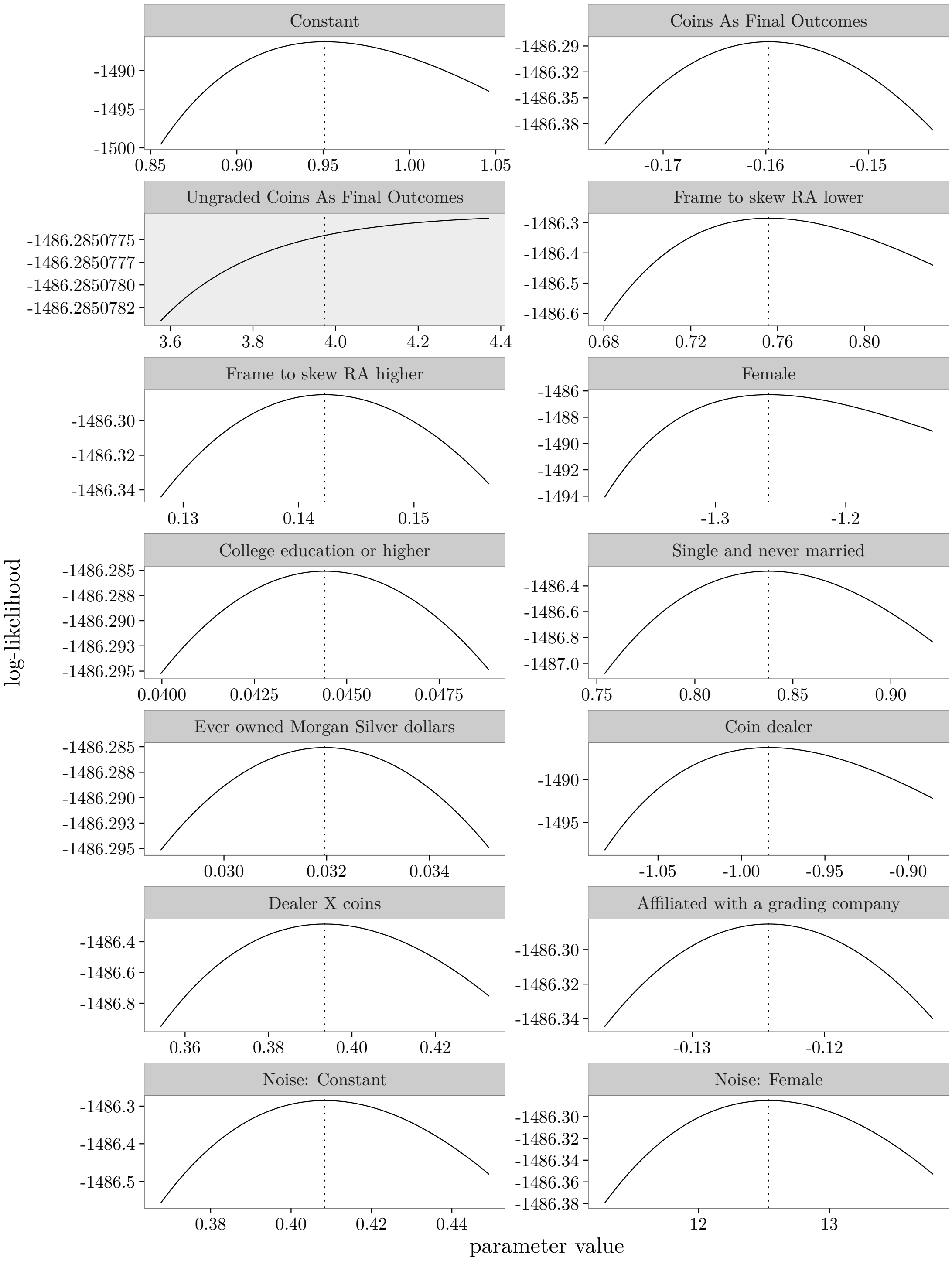
[3] In other words, the σs reported in Table II are the maximum likelihood estimates for the model

$$\Pr(\text{choice} = A) = \Phi\left(\frac{1}{100} \cdot \frac{E\left(\frac{m_A^{1-r}}{1-r}\right) - E\left(\frac{m_B^{1-r}}{1-r}\right)}{\sigma}\right)$$

and not for the model in Equation 1.1. For the corresponding Stata code, taken from HLT's supplementary material, see Listing 1.1 in Appendix 1.A1

[4] Other factors also play a role in producing this precise estimate. Using a different coding for the treatment indicators – an indicator for the *graded coins* treatment plus an indicator for the *ungraded coins* treatment instead of an indicator for both treatments involving coins and an indicator for *ungraded coins* – produces an *ungraded coins* coefficient of 257.354 even at the original tolerance setting. Moreover, for reasons detailed in Appendix 1.A5 even the sorting of the data set can significantly change the reported maximum.

[5] Interestingly, the near-complete flatness of the likelihood at the reported estimate is not reflected in the coefficient's standard error because standard errors are clustered on the subject level. Such clustering, which usually results in larger standard errors, can *shrink* standard errors if the contributions of individual observations to the score offset each other within a cluster. That this happens here is due to a sizable number of subjects whose responses are highly erratic (See Appendix 1.A2 for details). The non-robust standard error is 513.23.



Each sub-plot shows the log-likelihood for values of the respective parameter that are between 0.9 and 1.1 times the value shown in Table 1.1 while all other parameters are held fixed. The HLT estimate for each parameter is shown as a vertical dotted line.

Figure 1.1: Log-likelihood function in the neighborhood of the HLT parameter vector.

Perhaps surprisingly a treatment effect that is orders of magnitude or even infinitely larger than the one reported changes the model in no way that matters. All other coefficients are unaffected by the mis-estimated coefficient on the *ungraded coins* treatment dummy[6] and increasing the coefficient does not materially change the in-sample predictions of the model. The higher parameter values do, however, throw into sharp relief just what those predictions are even at the reported estimates.

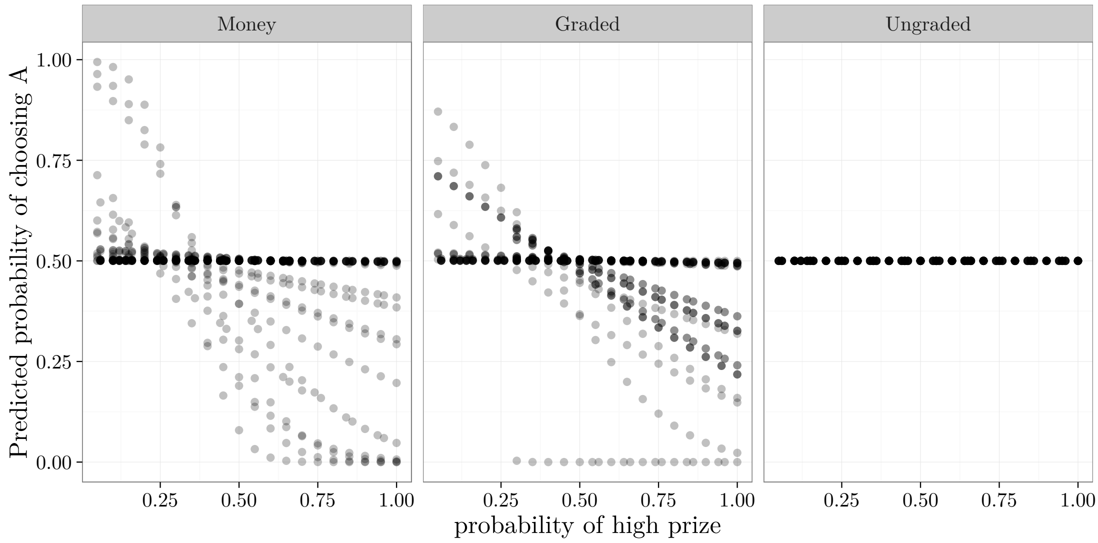
Consider the final row in the symmetric multiple price list. In this row subjects face the choice between a prize worth \$200 and a prize worth \$350, both paid with certainty. For any subject who values money this should be an easy choice. Indeed, the row is usually included in the list as a "sanity check" by which subjects who fail it can be excluded from subsequent analysis, if only as a robustness check.

Yet, assuming a CRRA coefficient of 4.78, the mean estimated CRRA coefficient in the ungraded coins treatment, the \$200 prize has an associated expected utility of $1 \cdot \frac{200^{1-4.78}}{1-4.78} = -5.304138 \times 10^{-10}$ and the \$350 prize has an associated expected utility of $1 \cdot \frac{350^{1-4.78}}{1-4.78} = -6.3963319 \times 10^{-11}$. Though the expected utility of the \$350 coin is higher the difference between these two numbers is so small even relative to the reported error standard deviation of 0.408 as to render this decision a virtual toss-up in which subjects randomize over the two options with roughly equal probability.

Under the correctly scaled error standard deviation the logic above holds even for subjects with substantially lower CRRA parameters. For the mean CRRA parameter in the *money* treatment of 0.88 the utility difference is 1.09 and for the mean CRRA parameter in the *graded coins* treatment of 0.76 the utility difference is 2.14. Both are small relative to an error standard deviation of 40.84 and so these subjects, too, randomize almost uniformly between \$350 and \$200.

Correspondingly, as Figure 1.2 shows, the number of subjects for which the model predicts toss-ups not just in the final decision but for *all decisions* in the list is substantial. For almost 75% of subjects in the *money* treatment and almost 60% of subjects in the *graded coins* treatment predicted choice probabilities for

[6] The other parameters are unchanged when the model is estimated with a lower tolerance. They are also unchanged when the model is estimated on the data from the *money* and *graded coins* treatments only. The data from the *ungraded coins* treatment, in other words, does nothing to identify any of these parameters



Points are semi-translucent to reduce overplotting.

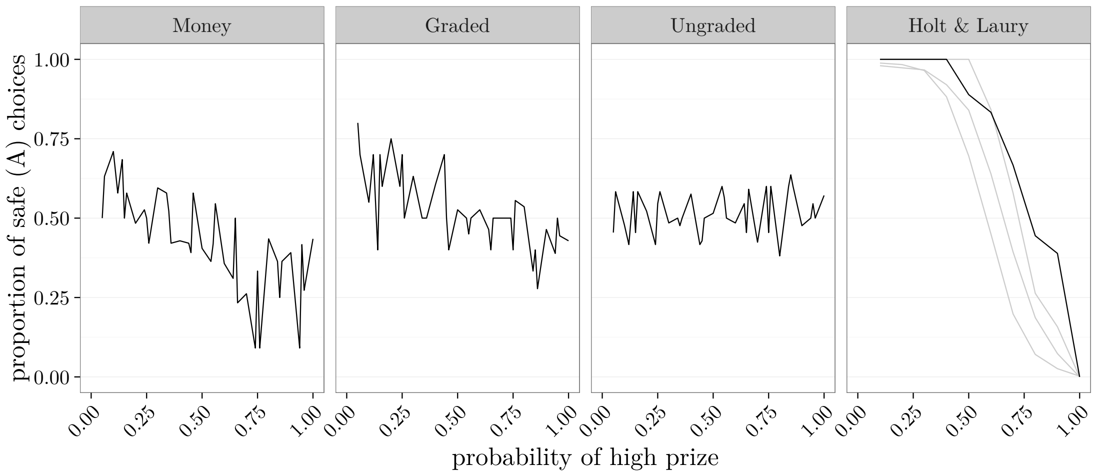
Figure 1.2: Choice probabilities for lottery A under the estimated model for all list items and treatments

the two lotteries never differ by more than 5 percentage points from 50:50 and for only a handful of subjects are predicted choice probabilities ever close to zero or one.

This is most extreme in the treatment that identifies the effect of background risk. In the *ungraded coins* treatment, the model predicts a toss-up for *all* decisions and *all* subjects. The estimated treatment effect on the CRRA parameter for the *ungraded coins* treatment is so high, in other words, that the structural component of the model – expected utility – has virtually no explanatory power for the data.[7] A higher coefficient for the *ungraded coins* treatment dummy would only reduce the explanatory power further.

These conclusions naturally raise the question how the estimates come to be,

[7] The value of the parameter is also so high that none of the multiple price lists subjects were presented with could have identified it had choice been deterministic. Three different multiple price lists were used in the experiment, one in which the probability on the high prize varied from 0.05 to 1 in step sizes of 0.05, and two in which the lists were skewed towards low and high probabilities to test for an anchoring effect or centrality bias. As shown in HLT's Figure 3 the identification regions for these three lists – the lowest and highest CRRA parameters which would produce a switching point in the list – were [-2.03,2.61], [-1.85,1.10] and [-0.23,2.78] respectively.



The plotted Holt&Laury data are (from left to right) from the 1x, 20x (N = 93), 50x (N = 19) and 90x (N = 18) real money treatments. The most closely comparable 90x treatment in black, treatments with weaker incentives in light gray.

Figure 1.3: Proportion of A choices by treatment

a question that is best answered by looking at the raw data. Figure 1.3 shows the proportion of subjects who choose the safe lottery A as the probability p of receiving the lotteries' high prize varies from 0.05 to 1. For comparison the plot also shows the proportion of safe choices from the incentivized treatments in Holt and Laury (2002) with the most closely comparable treatment in black and the treatments with lower stakes in gray[8]. If the multiple price list works as designed this probability should (if subjects are not extremely risk loving) be one for $p = 0.05$ and then monotonically decline to zero for $p = 1$, i.e. when subjects face the choice between certain \$200 and \$350. Yet, across all three main treatments the proportion of A choices is never above 0.8 and never below 0.09. Moreover, the proportion does not fall monotonically as in Holt and Laury but oscillates, the consequence of a large number of subjects who switch between options A and B not once but repeatedly. A look at the individual data — shown in full in Appendix 1.A2 — also suggests that these switches may not be as unsystematic

[8] Prizes in the most closely comparable, Holt and Laury's 90x, real stakes treatment, were \$100 and \$80 for the safe lottery, and \$192.50 and \$5 for the risky lottery. As is clearly visible in Figure 1.3, in treatments with lower stakes as well (as those with hypothetical choices), subjects choose the safe option less often in almost all decisions. However, in all treatments the number of safe choices is close to 100% for the lowest probability on the high prize and close to 0% for the highest probability on the high prize.

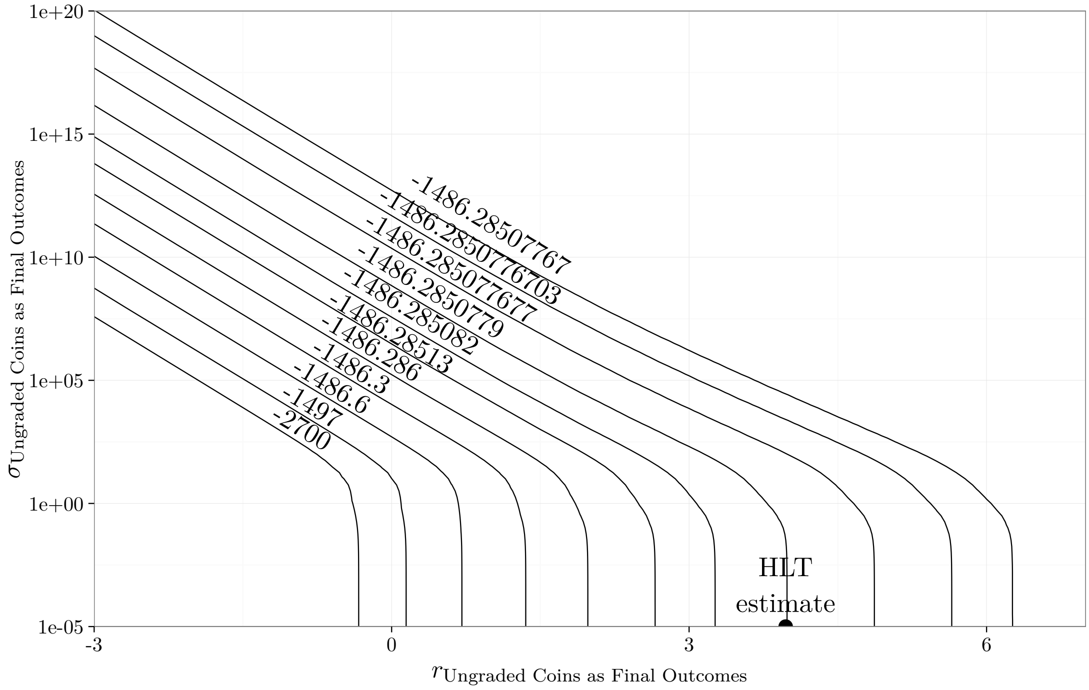
as the model assumes. Instead, the responses of a substantial number of subjects look patterned, with subjects alternating between A and B or repeating other sequences.

Importantly, while the proportion of A choices decreases in p for the *money* and *graded coins* treatments, it is essentially flat (up to 'error') for the *ungraded coins* treatment. Changing the difference in expected values between the risky and the safe lottery from $\$-73.25$ to $150 has no systematic effect on the frequencies with which the two lotteries are chosen!

The key to understanding how these data turn into the reported estimates lies in the mechanics of the stochastic choice model. When holding the standard deviation of the choice error σ fixed increasing the CRRA parameter has more than one effect: First, as one would expect, the utility function becomes more concave, which moves the point at which the difference in utilities between lottery A and B turns negative further down the list. Second, the absolute scale of utilities decreases, which makes choices more random.

The only way the model as it is specified can fit the complete absence of a response to such large changes in the expected values of the gambles is to make the CRRA coefficient so large as to make the absolute scale of the utility function minuscule. Given this scale the structural part of the model becomes completely irrelevant for choice. The extreme risk aversion HLT report for subjects exposed to *ungraded coins* is a very peculiar kind of risk aversion in which expected utility commands subjects to choose the safe option in all but the final row, but in which they err so much that the choices they make are indistinguishable from being made entirely at random.

Another way of making the same point fully within HLT's model structure is to consider an alternative specification in which replacing graded with ungraded coins can have an effect not only on subjects' CRRA parameter but also on the magnitude of the error standard deviation σ_i. Simply adding an indicator variable to the equation determining the error standard deviation yields an effect of ungraded coins on noise of 9862.62 and an effect on the CRRA parameter of 4.581×10^{12} (all other coefficients are essentially the same as in Table 1.1). These estimates, however, are misleading for this model does not even appear to be identified. Figure 1.4 shows the likelihood contour for the model in which the treatment effects



All parameters not shown in the graph are held at the values they obtain after an indicator for *ungraded coins* is added to HLT's specification and the model is estimated to a tolerance on the scaled gradient of 1e-9 (see Table 1.A1 in Appendix 1.A3). Log-likelihoods are computed on a grid with $r_{\text{Ungraded Coins as Final Outcomes}} \in \{-3, -2.9, \ldots, 7.9, 8\}$ and $\sigma_{\text{Ungraded Coins as Final Outcomes}} \in \{10^{-5}, 10^{-4.5}, \ldots, 10^{19.5}, 10^{20}\}$ and then interpolated. Notches on highest contour line are artifacts of interpolation. Point shows original HLT estimate.

Figure 1.4: Likelihood contour for treatment effects of ungraded coins on both CRRA coefficient $r_{\text{Ungraded Coins as Final Outcomes}}$ and error standard deviation $\sigma_{\text{Ungraded Coins as Final Outcomes}}$

of *ungraded coins* on both CRRA and noise standard deviation are varied while all other parameters are held fixed at HLT's estimates. The figure shows that the two parameters can be traded off against each other, seemingly at will. HLT's original estimate for the treatment effect on the CRRA coefficient is shown in the lower right but the identical likelihood is reached with a null effect on the CRRA coefficient and a suitably large effect on the noise parameter. Identification of the treatment effect on the CRRA parameter in HLT therefore depends crucially on the assumption of no treatment effect on the magnitude of the choice error parameter. Such a lack of a treatment effect on the magnitude of the choice error parameter seems hard to justify both because of the link between risk aversion and noisiness that is baked into the model and the fact that the ungraded coins treatment may very well have been more difficult for subjects to understand or required more effort on their part.[9]

1.2.2 Variable Selection

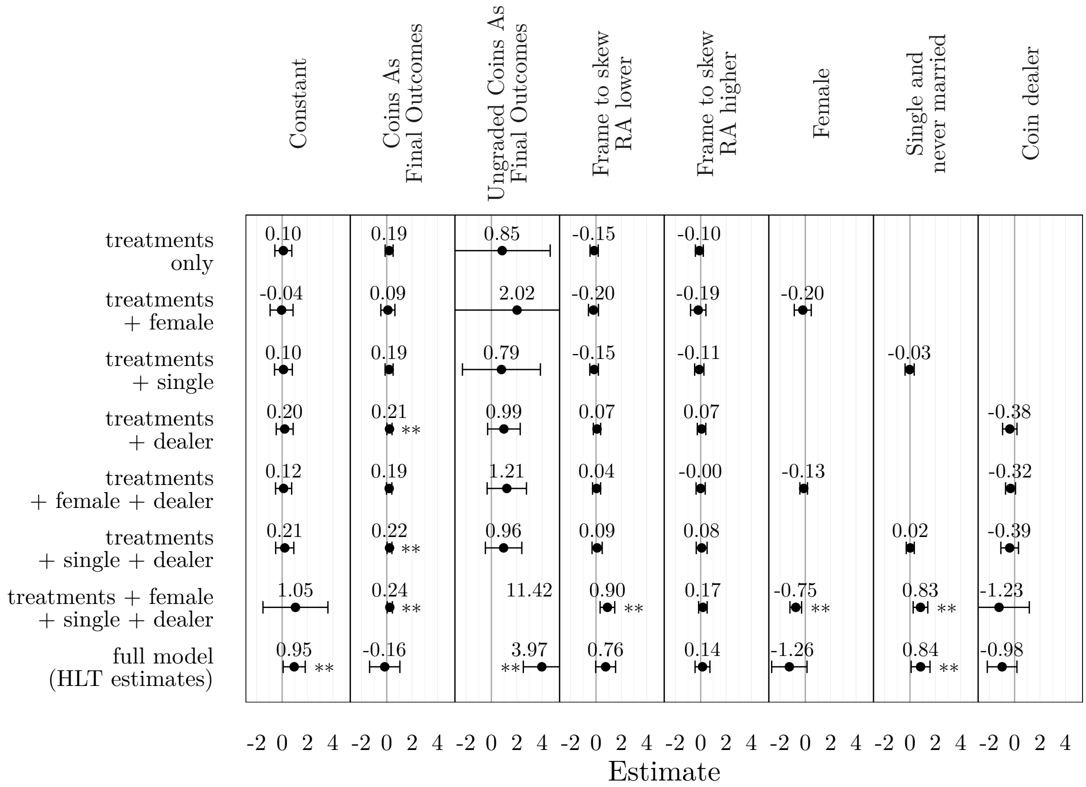
As was shown in the last section, HLT's enormous treatment effect for *Ungraded Coins as Final Outcomes* is the result of experimental data in which there is no behavioral response to changes in stimuli and a stochastic choice model that, in the absence of a possibility to attribute it to choice errors, takes this absence to be evidence of extreme risk aversion. Results for the other treatment effects, while identified by data from treatments in which subjects do, however noisily, respond to incentives, are dependent on model specification and, in particular, on the set of controls for individual characteristics included in the model.

Figure 1.5 shows point estimates and 95% confidence intervals for the four treatment effects and the effects of three of the seven individual characteristics included in HLT's original specification —– *Female*, *Coin dealer* and *Single and never married*. It does so for specifications that contain the four treatment indicators and any possible combination of the three individual characteristics, for a total of seven specifications. For comparison, the Figure also shows estimates from HLT's original specification.

Note that in what one might consider a minimal specification, one that contains only the four treatment indicators, all estimated effects are small and none are statistically significantly different from zero. In fact, for this model an omnibus F-test cannot reject the null ($\chi^2(4) = 2.79$, $p = 0.59$).

Only after the addition of controls for demographic characteristics does this change. It is, of course, entirely correct to control for individual characteristics. This not being a linear model, the estimated treatment effects would otherwise suffer from omitted variable bias despite being orthogonal by design. *Which* controls to include, however, is not an innocuous choice. Even within the small subset

[9] In the *money* treatment subjects were told the monetary value of the four prizes. In the *graded coins* treatment subjects were presented with four coins as prizes, all of which had an attached certificate attesting to the coin's condition. In the *ungraded coins* treatment the same four coins were used. The certificates, however, had been removed by the experimenters. The coins would therefore have looked identical unless examined.



Confidence intervals based on clustered standard errors. *Noise: Constant* not shown for all models. College education or higher, *Ever owned Morgan Silver dollars*, *Dealer × coins*, *Affiliated with a grading company*, *Noise: Constant* and *Noise: Female* not shown for full model. "treatments + female + single" now shown because the specification shows the pathology described in Section 1.2.2
**: $p < 0.05$

Figure 1.5: Point estimates and 95% confidence intervals for models involving treatment indicators and the three individual characteristics *Female*, *Single and never married* and *Coin dealer*

of possible specifications shown in Figure 1.5, the statistical significance, magnitude and even sign of all coefficient except for the one on *Ungraded Coins as Final Outcomes* is dependent on the specification.

Including all three individual characteristics in the model —– call this the reduced specification — yields estimates that are broadly similar to those in HLT's original specification and a log-likelihood that is only 4.10 points lower[10]. Esti-

[10] It is unclear how these two models ought to be formally compared. A Wald test for a joint restriction on HLT's original specification at HLT's parameter vector cannot reject the null ($\chi^2(5) = 2.31$, $p = 0.80$). However, this test assumes that HLT's parameter vector is likelihood maximal, which it is not, see below.

mated parameters for the individual characteristics that remain in the model are very similar to those in the original specification. The picture for the treatment effects, however, looks very different. Relative to the original estimates, the treatment effect of *Coins as Final Outcomes* reverses sign and becomes statistically significantly different from zero (point estimate: 0.24, $p = 0.02$). Under this specification, in other words, using "natural" instead of artificial prizes has a statistically significant influence on elicited risk preferences. The estimated treatment effect of *Ungraded Coins as Final Outcomes* is qualitatively similar to that in the original specification, that is, it continues to suffer from the issue detailed in the last section. Lastly, in HLT's original specification the cross-treatment in which probabilities in the multiple price lists are skewed towards values that would produce lower measured risk aversion if subjects had a tendency to switch in the middle of the list is estimated to *raise* the CRRA coefficient by 0.756 but is not statistically significantly different from zero at the 5% level. In the reduced specification, the treatment effect is estimated to be 0.90 and is now highly statistically significant ($p = 0.002$).

What about specifications with other sets of demographic controls? It is at this point that one must confront the unfortunate reality that the model is not identified for most of the specifications that contain more than three individual characteristics. Table 1.A2 in the Appendix shows the results of numerical maximization for specifications involving all possible combinations of the seven individual characteristics included in HLT's original specification.[11] For many of the specifications the model shows groups of parameters moving towards infinity in opposite directions.

This phenomenon is similar to the (quasi-)complete separation which is sometimes encountered in standard binary choice models when a combination of variables allows the model to predict the dependent variable perfectly for a subset of observations. In such cases the likelihood does not have a maximum because predicted choice probabilities under the model can be driven ever closer to zero

[11] HLT's original specification does not use all variables available in the dataset. The dataset also contains a number of other variables: An experimenter effect, information about the years of experience in the coin and paper money market, the number of shows attended and the number of coins graded in a year, age, more fine-grained educational attainment, income level, the marital status, size of the household, information about whether and how much a subject smokes, which day of the 3-day show the experiment was conducted on, and whether a participant only deals in graded or ungraded coins.

or one by choosing more extreme parameter values. Similarly, adding individual characteristics to HLT's model sometimes gives it enough flexibility to drive the coefficient of risk aversion of a group of subjects who are identical on the included variables either towards positive infinity, which drives choice probabilities under the model towards 0.5, or towards negative infinity, which drives choice probabilities for the riskier option B to one. In both case no likelihood maximum can exist because choices can always be moved closer to 0.5 by increasing the CRRA coefficient or closer to 1 by decreasing it.

HLT's data contain observations of subjects who choose the risky option B throughout and subjects who choose very unsystematically. For the former group the model can gain likelihood by assigning to the group a negative CRRA coefficient while for the latter group it can gain likelihood by assigning the group a very large, positive CRRA coefficient. As long as the specification is relatively sparse each cell in the partition is unlikely to contain only subjects who make such "extreme" choices and estimated coefficients and CRRA parameters will be "reasonable". Adding controls for individual characteristics, however, allows the models to partition subjects ever more finely and on HLT's data the partition soon becomes fine enough to isolate extreme subjects. Once this happens, the model can drive the coefficient of one of the individual characteristics shared by these subjects towards infinity in either direction. The coefficients on some of the other characteristics, meanwhile, move in the opposite direction so the CRRA coefficient of subjects who share some but not all the characteristic does not also move.

On HLT's data such estimates are common and the model is easily pushed towards them. Starting from the minimal specification that contains only treatment indicators and in which the choice error does not differ by gender adding just *one* demographic variable can already produce extreme results. By the time the model contains four individual characteristics only 5 of the 35 specification still have well-defined likelihood maxima. By five individual characteristics, these results are universal (for details, see Appendix Section 1.A4).

The same problem ails HLT's original specification: HLT use a Newton-Raphson algorithm to find the estimates reported in the paper, at a log-likelihood of -1486.285. All other maximization algorithms offered by Stata do not converge on a likelihood maximum. For the BHHH algorithm, however, intermediate solution candidates achieve log-likelihoods above -1460. Generalized simulated annealing, which explores the parameter space randomly and therefore does not rely on the problem being globally concave finds a solution that is better still (log-likelihood: -1438.880). This draws into question the original estimates and inference based upon them in their entirety.

Luckily, the reduced specification that contains only three of the individual characteristics does not suffer from this pathology. For it, the likelihood maximum seems to be well-defined aside from the issue with the *Ungraded Coins as Final Outcomes* coefficient discussed in Section 1.2.1.[12]

All in all, the positive treatment effect of using *Coins as Final Outcomes* in the reduced specification is not anomalous. In fact, it's HLT's estimate of a negative treatment effect that is. It is positive and of roughly equal magnitude for all specifications shown in Figure 1.5 and of the 64 specifications in which estimates do not show signs of the pathology described above, none feature a negative point estimate for *Coins as Final Outcomes*. Its statistical significance, however, does depend on which set of individual characteristics the model contains. The same is true for the effect of the treatment designed to lower measured risk aversion by skewing the multiple price list. This effect was already sizable in HLT's specification but not statistically different from zero. In the reduced specification the effect is similarly large and statistically significant, but this is the only one of the models shown in Figure 1.5 for which either are true. In most the effect is of variable sign, small and not statistically significant.

1.3 Discussion and Conclusion

Harrison, List, and Towe raise important methodological questions and use an elegant experimental design to attempt to answer them. Unfortunately, when probed the experimental responses simply do not appear to be up to the task of providing

[12] While the reduced specification has a well-defined maximum on this sample, the same is not always true for the model estimated on resamples of the data, e.g. for the purposes of bootstrapping standard errors. Indeed, not even the minimal specification that contains only treatment indicators seems to reliably possess a likelihood maximum under re-sampling

answers. A majority of subjects makes choices that deviate from expected utility maximization, which must mean that the results of any econometric analysis will strongly depend on whether such “errant” observations are retained in the sample, on the exact form deviations from rationality are assumed to take if they are and on the modelling of individual heterogeneity. Wilcox (2011) has forcefully made this point theoretically while Andersson, Holm, Tyran, and Wengström (2013) have recently provided another empirical illustration: the oft-reported negative relationship between risk aversion and cognitive ability may also be a statistical artifact in which a higher error rate among subjects with low cognitive ability is (mis-)identified by the econometric model as higher risk aversion.

The search for appropriate stochastic identifying assumptions will likely continue for some time. In the absence of widely agreed-upon methods it is important to assess the sensitivity of model estimates to alternative identifying assumptions and to handle the interpretation of estimates from models which imperfectly separate structure from noise with great care.

1.4 References

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation Using Multiple Price List Formats. *Experimental Economics*, *9*(4), 383–405.

Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2013). Risk Aversion Relates to Cognitive Ability: Fact or Fiction? *IFN Working Paper*, *964*.

Charness, G. & Viceisza, A. (2012). *Comprehension and Risk Elicitation in the Field: Evidence from Rural Senegal.*

Filippin, A. & Crosetto, P. (2016). A Reconsideration of Gender Differences in Risk Attitudes. *Management Science*, *forthcomin*.

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, *80*(1), 27–38.

Gollier, C. & Pratt, J. (1996). Risk vulnerability and the tempering effect of background risk. *Econometrica*, *64*(5), 1109–1123.

Harrison, G. W., List, J. A., & Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: a case study of risk aversion. *Econometrica*, *75*(2), 433–458.

Harrison, G. W. & Rutström, E. E. (2008). Risk Aversion in the Laboratory. In J. C. Cox & G. W. Harrison (Eds.), *Risk aversion in experiments, research in experimental economics* (Vol. 12, pp. 41–196). Risk Aversion in Experiments. Emerald Group Publishing.

Hey, J. & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*(6), 1291–1326.

Holt, C. & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Kirkpatrick, S., Gelatt, C. D., P., V. M., & Vecchi, M. (1983). Optimization by Simulated Annealing. *Science*, *220*(4598), 671–680.

Lévy-Garboua, L., Maafi, H., Masclet, D., & Terracol, A. (2011). Risk Aversion and Framing Effects. *Experimental Economics*, *15*(1), 128–144.

Nash, J. C. & Varadhan, R. (2011). Unifying Optimization Algorithms to Aid Software System Users: optimx for R. *Journal of Statistical Software*, *43*(9), 1–14.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison. In J. C. Cox & G. W. Harrison (Eds.), *Risk aversion in experiments, research in experimental economics* (Vol. 12, pp. 197–292). Research in Experimental Economics. Emerald Group Publishing Limited.

Wilcox, N. T. (2011). 'stochastically more risk averse:' a contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, *162*(1), 89–104.

Yang Xiang, Gubian, S., Suomela, B., & Hoeng, J. (2013). Generalized Simulated Annealing for Efficient Global Optimization: the GenSA Package for R. *The R Journal*, *5*(1), 13–28.

1.5 Appendices

1.A1 Original Code

```
* code up an EUT likelihood function for CRRA with Fechner errors
program define MLcrra_f

    args lnf r noise
    tempvar prob1l prob2l prob1r prob2r y1l y2l y1r y2r euL euR euDiff scale

    quietly {

        generate double `prob1l' = $ML_y2
        generate double `prob2l' = $ML_y3

        generate double `prob1r' = $ML_y4
        generate double `prob2r' = $ML_y5

        generate double ‘scale’ = 1/100

        generate double `y1l' = `scale'*((($ML_y10+$ML_y6)^(1-`r'))/(1-`r'))
        generate double `y2l' = `scale'*((($ML_y10+$ML_y7)^(1-`r'))/(1-`r'))
        generate double `y1r' = `scale'*((($ML_y10+$ML_y8)^(1-`r'))/(1-`r'))
        generate double `y2r' = `scale'*((($ML_y10+$ML_y9)^(1-`r'))/(1-`r'))

        generate double `euL' = (`prob1l'*`y1l')+(`prob2l'*`y2l')
        generate double `euR' = (`prob1r'*`y1r')+(`prob2r'*`y2r')

        generate double `euDiff' = (`euR'-`euL')/`noise'

        replace `lnf' = ln($cdf( `euDiff')) if $ML_y1==1
        replace `lnf' = ln($cdf(-`euDiff')) if $ML_y1==0

    }

end
```

Listing 1.1: Code defining the stochastic choice model, taken from HLT's supplementary material. Excerpt from coins_ml_code.do, lines 63 – 94. Line that scales down the utility difference by a factor of $\frac{1}{100}$ highlighted in yellow.

1.A2 Individual experimental responses

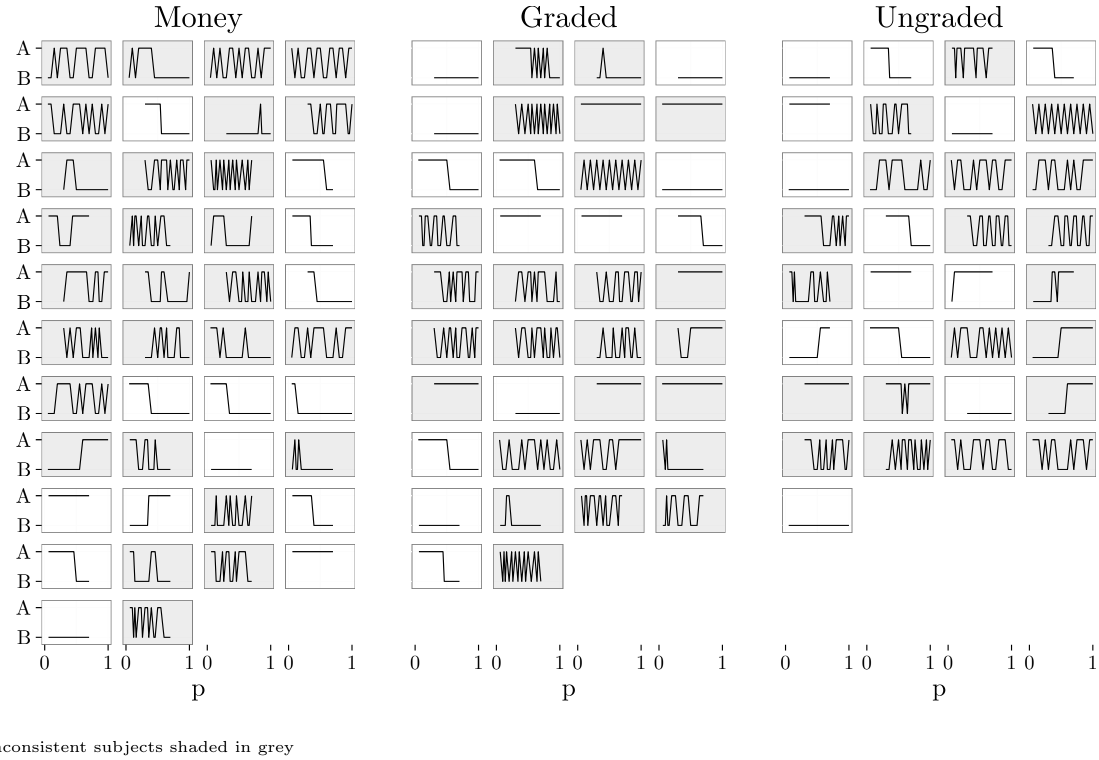


Figure 1.A1: Individual responses to the multiple price list

Figure 1.A1 plots the individual choice data. What is remarkable about these data is that the share of subjects who switch multiple times is extremely high. 56% of subjects switch between A and B repeatedly and often do so in a patterned fashion.

Two of the three alternative multiple price lists that were used also contain a second "sanity check". In the last item of the list subjects face the choice between the high prize of lottery A and the high prize of lottery B *with certainty*. Because this high prize is higher in lottery B than in lottery A every subject with an increasing utility function must choose B. Yet, of the 72 subjects who face such a choice 47% choose lottery A, a lottery that is dominated not just stochastically but deterministically.

All in all, of the 113 subjects a full 65% make choices that violate first-order stochastic dominance. Compared with other studies that use Holt and Laury-style lists this is extremely high. Holt and Laury (2002) report inconsistent responses for 20% of subjects in their low-stakes condition and 13% in the high-stakes condition. All treatments in Holt and Laury use smaller monetary prizes than HLT. In a survey of all published papers which use such lists Filippin and Crosetto (2016) find that the percentage of inconsistent subject is 13.9% across all studies. The incidence of inconsistencies varies predictably with several aspects of the list (see e.g. Lévy-Garboua, Maafi, Masclet, & Terracol, 2011) and with the subject pool used, with university subjects displaying fewer inconsistencies and subjects outside of the lab producing rates of inconsistency that are often above 50% (see Charness & Viceisza, 2012, and references therein).

While the proportion of inconsistent subjects if often so small that estimating the model with or without them makes very little difference for the results, this is not the case for HLT's sample. Estimating the main model only on those subjects who satisfy strict rationality drastically shrinks the treatment effect of ungraded coins and makes all treatment effects statistically insignificant (see Table 1.A1).[13]

[13] Note, however, that eliminating responses that are inconsistent with strict rationality is not a neutral criterion in the presence of subjects who choose naively. A subject who naively chooses A throughout would be eliminated from the sample for choosing the dominated option in the last row, while a subject who chooses B throughout would be retained and interpreted as extremely risk loving, a fact that may explain the negative estimate for the constant in the model.

1.A3 Alternative Estimation Results

	Parameter	Original Model	Lower Tolerance	Without *ungraded coins* treatment	Original + $\sigma_{\text{Ungraded Coins}}$	Reduced Specification	No Inconsistent Subjects
r	Constant	0.951	0.951	0.951	0.951	1.050	-0.394
		(0.444)	(0.444)	(0.445)	(0.444)	(1.306)	(1.687)
	Coins As Final Outcomes	-0.160	-0.160	-0.160	-0.160	0.240	-0.124
		(0.610)	(0.610)	(0.611)	(0.610)	(0.104)	(0.238)
	Ungraded Coins As Final Outcomes	3.974	316.638		>1000	11.423	0.169
		(0.744)	.		.	.	(0.133)
	Frame to skew RA lower	0.756	0.756	0.756	0.756	0.897	0.294
		(0.399)	(0.399)	(0.400)	(0.399)	(0.294)	(0.527)
	Frame to skew RA higher	0.142	0.142	0.142	0.142	0.174	0.074
		(0.299)	(0.299)	(0.299)	(0.299)	(0.172)	(0.396)
	Female	-1.259	-1.259	-1.259	-1.259	-0.753	-0.999
		(0.714)	(0.712)	(0.710)	(0.711)	(0.235)	(3.575)
	College education or higher	0.044	0.044	0.044	0.044		0.141
		(0.209)	(0.209)	(0.210)	(0.209)		(0.335)
	Single and never married	0.838	0.838	0.838	0.838	0.829	-0.107
		(0.380)	(0.379)	(0.380)	(0.379)	(0.294)	(0.233)
	Ever owned Morgan Silver dollars	0.032	0.032	0.032	0.032		0.092
		(0.573)	(0.573)	(0.574)	(0.573)		(0.335)
	Coin dealer	-0.984	-0.984	-0.984	-0.984	-1.227	-0.226
		(0.596)	(0.596)	(0.597)	(0.596)	(1.217)	(0.145)
	Dealer X coins	0.394	0.394	0.394	0.394		0.363
		(0.472)	(0.472)	(0.473)	(0.472)		(0.387)
	Affiliated with a grading company	-0.124	-0.124	-0.124	-0.124		0.051
		(0.283)	(0.283)	(0.284)	(0.283)		(0.416)
σ	Constant	0.408	0.408	0.408	0.408	0.979	1.868
		(0.734)	(0.734)	(0.735)	(0.734)	(1.385)	(10.940)
	Female	12.540	12.543	12.542	12.539		304.166
		(38.505)	(38.335)	(38.055)	(38.291)		(7234.324)
	Ungraded Coins As Final Outcomes				>1000		
					.		
LL		-1486.2850779	-1486.2850777	-1028.8079385	-1486.2850777	-1490.387084	-387.476626
N		112	112	79	112	112	39

Clustered standard errors in parentheses. "." = standard errors not reported by Stata.
For comparability all coefficients for σ reported on the same scale as in HLT.
"Original Model" refers to the estimates reported in HLT. "Lower Tolerance" refers to the same specification but estimated to a tolerance of 10^{-8} on the scaled gradient (Stata option `nrtolerance(1e-8)`). "Without *ungraded coins* treatment" refers to the Original Model estimated on a sample without subjects in the *ungraded coins* treatment. "Original + $\sigma_{\text{Ungraded Coins}}$" refers to a specification that contains not only a treatment effect of *ungraded coins* on the CRRA parameter but also one on the noise parameter σ. "Reduced specification" refers to a specification in which *College education or higher*, *Ever owned Morgan Silver dollars*, *Dealer* $\times$ *coins*, *Affiliated with a grading company* and the gender effect on the noise parameter have been removed from the model. "No Inconsistent Subjects" refers to "Original Model" estimated on a reduced data set that contains only subjects who switch at most once in the multiple price list and choose $350 over $200 in the final row of the list.

Table 1.A1: Alternative estimation results

1.A4 Point estimates for all specifications

Table 1.A2 contains the results of numerical maximization for all model specifications in which the CRRA parameter r is determined by a constant, the four treatment indicators *Coins As Final Outcomes*, *Ungraded Coins As Final Outcomes*,

Frame to skew RA lower and *Frame to skew RA higher*, any possible combination of the seven individual characteristics contained in HLT's original model, and the noise parameter σ is either identical across all subjects or varies by gender. This yields a total of 256 specifications.

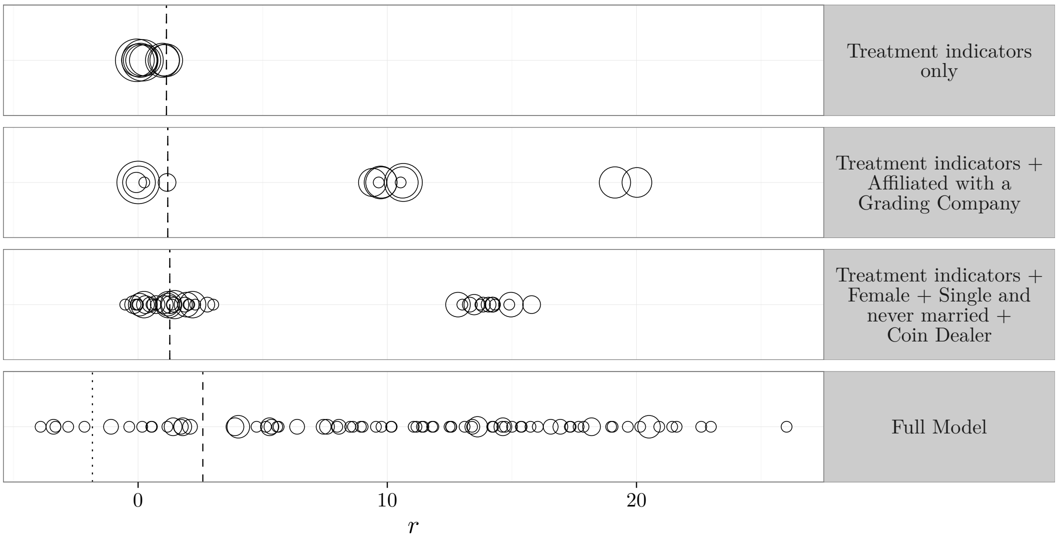
I use the term "results of numerical maximization" rather than "maximum likelihood estimates" advisedly as the likelihood functions for many of these specifications, including HLT's original specification, can be extremely rugged, with several local maxima in which standard hill-climbing algorithms can easily get stuck. Even maximization methods which do not rely on the likelihood being globally concave and explore the parameter space randomly do not reliably find the likelihood maximum and it is nigh impossible to know whether the parameter vectors found in extensive numerical searches and displayed in Table 1.A2 really do represent the global maxima if they exist. Even specifications that include far fewer variables than HLT's original specification, however, very often do considerably better than HLT's original estimates.

I use the term "results of numerical maximization" also because for many specifications the log-likelihood does not appear to possess a maximum at all for the reason discussed in Section 1.2.2.

As an example of the model driving predictions to extremes after the addition of an individual characteristic, start with the minimal specification that contains only treatment indicators and no individual characteristics. This is a model that is well-behaved on HLT's dataset: Essentially all maximization algorithms reliably find the same likelihood maximizer, which produces the distribution of CRRA coefficients shown in the top panel of Figure 1.A2.

Adding only *Affiliated with a grading company* to this model, however, changes the parameter vector found via numerical maximization significantly: All coefficients but the one associated with *Ungraded Coins as Final Outcomes* are either sharply negative or sharply positive for a distribution of CRRA coefficients that looks like the one shown in the second panel of Figure 1.A2.

The third panel of Figure 1.A2 shows the distribution of CRRA coefficients in the reduced specification that contains treatment indicators, *Female*, *Single and never married* and *Coin Dealer*. In this specification subjects in the *ungraded coins* treatment have very high CRRA coefficients while subjects in the two other



"Full Model" is HLT's specification evaluated at the parameter vector found via simulated annealing (see Table 1.A2). Area of circles proportional to the number of subjects. For context, vertical lines mark the CRRA coefficient for which the probability of choosing the dominant \$350 over the dominated \$200 option in the last row of the multiple price list is exactly 0.55, i.e. barely better than chance, given the estimated noise standard deviation σ. In "Full Model", the dashed line marks this threshold for male subjects and the dotted line marks it for female subjects. In all other models, the dashed line marks the threshold for all subjects.

Figure 1.A2: Distributions of CRRA coefficients r for all subjects for four specifications at parameter values shown in Table 1.A2

treatments have much lower CRRA coefficients[14]. The fourth panel shows the distribution of CRRA coefficients for HLT's full model evaluated at the parameter vector found via generalized simulated annealing. The full model contains enough controls for individual characteristics to allow the model to finely differentiate between different groups of subjects. Note, however, that for almost all subjects the parameters imply essentially random choice.

Of all 256 specifications shown in Table 1.A2 at least 192 show evidence of this pathology[15]

The literature offers two main solutions to this problem:

1. Removing covariates until the maximization problem becomes well-defined again

[14] Note that even comparatively small coefficients often imply choices that do not differ much from complete randomization.

[15] based on the ad hoc criterion that among the estimates for all r parameters other than the one on *Ungraded Coins as Final Outcomes* there is at least one coefficient which is larger than 3 and one coefficient that is smaller than -3.

2. Penalized likelihood estimation (Firth, 1993), a procedure in which a term that penalizes extreme parameter values is added to the log-likelihood. This ensures that the entire maximization problem is guaranteed to have a global maximum even if the likelihood function does not.

Neither of these solutions is ideal. In as far as covariates help predict the dependent variable their removal may restore identification but introduce bias at the same time. Penalized likelihood estimation, on the other hand, will ensure the existence of a maximum but do so by essentially imposing specific values on groups of parameters (how far the parameters are allowed to move away from each other will depend on how high the penalty for large parameter values is), some of which are the treatment indicators which are of main inferential interest.

Likelihood maximization — Methodology

The estimates displayed in Table 1.A2 are the results of extensive efforts to maximize the log-likelihood functions associated with all 256 specifications. Numerical maximization was performed using a number of different approaches that differ in the maximization algorithm used and in the way the algorithms are initialized.

First, I use a variety of maximization algorithms that break down into two classes:

- **hill-climbing methods**: These include simplex methods like Nelder-Mead, gradient ascent methods like BFGS, CG (conjugate gradient), nlminb (a gradient ascent method using PORT routines), spg (spectral projected gradient) and ucminf, and derivative-free methods using trust regions like newuoa, bobyqa, and using other techniques like the Hooke-Jeeves algorithm for derivative-free optimization. All methods are provided by R's `optimx` package (Nash & Varadhan, 2011). These algorithms are generally fast to converge to an optimum but may get stuck in local optima if the likelihood function is not globally concave, in which case the value at which the parameter vector is initialized may matter.[16]
- **generalized simulated annealing** as implemented by R's `GenSA` package (Yang Xiang, Gubian, Suomela, & Hoeng, 2013). Simulated annealing (Kirk-

[16] Implementation details: All algorithms were run for a maximum of 2500 iterations with default criteria and threshold for convergence.

patrick, Gelatt, P., & Vecchi, 1983) is a method that takes random jumps through the parameter space in search of higher values of the objective function and in which the size of these jumps is reduced over time. This method is generally more robust to getting stuck in local optima but is very slow.[17]

Second, to further reduce the likelihood of getting stuck in local optima or to miss optima, each of the above maximization algorithms were initialized in two different ways:

- **independent**: All parameters in the r equation were set to 0 and all parameters in the noise equation were set to 2.72 (e).[18]
- **iterative**: For each specification, the results from all nested specifications were gathered and from this set the result with the highest log-likelihood was selected. The maximization algorithm was then initialized at this parameter vector while setting the parameters not contained in the nested model to zero.

For each of the 256 specifications, Table 1.A2 shows the parameter vector which achieves the highest log-likelihood among all maximization algorithm and among all initialization procedures, no matter whether the algorithm in question successfully converged.

[17] Implementation details: the algorithm searched for optima in a hypercube, with all r-parameters in the interval $[-15, 15]$ and all σ-parameters in the interval $[\exp(-20), \exp(20)]$. The algorithm was run for a maximum of $100,000$ iterations and would stop if no improvement had occurred for $5,000$ steps.

[18] where, again, noise parameters are on the same scale as reported in HLT for comparability

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
TI only		0.10	0.19	0.85	-0.15	-0.10								2.05		-1534.74	2259
female		-0.04	0.09	2.02	-0.20	-0.19	-0.20							11.19		-1514.48	2239
college		0.04	0.19	1.06	-0.17	-0.12		0.04						2.68		-1534.46	2259
single		0.10	0.19	0.79	-0.15	-0.11			-0.03					2.06		-1534.68	2259
ms		0.09	0.19	0.86	-0.15	-0.10				0.01				2.05		-1534.74	2259
dealer		0.20	0.21	0.99	0.07	0.07					-0.38			1.38		-1511.45	2259
dealer.coins		0.10	0.55	0.86	-0.09	-0.02						-0.56		1.49		-1527.30	2259
grcomp		9.41	10.61	-0.88	-9.41	-9.37							-9.48	1.54		-1530.47	2259
female, college		14.96	-3.48	4.28	-5.69	-7.13	-4.59	-3.06						0.01		-1506.29	2239
female, single		14.73	-13.38	1.06	6.00	14.41	13.18		-7.05					0.00		-1501.25	2239
female, ms		1.47	0.48	2.12	-0.17	-0.99	-1.07			-0.94				38.60		-1503.15	2239
female, dealer		0.12	0.19	1.21	0.04	-0.00	-0.13				-0.32			2.45		-1496.08	2239
female, dealer.coins		-0.06	0.30	1.20	-0.14	-0.07	-0.16					-0.37		6.80		-1508.36	2239
female, grcomp		15.00	-2.87	4.42	-2.30	-5.38	-9.56						-5.59	0.91		-1502.84	2239
college, single		14.83	-9.59	-0.13	3.63	6.63		11.43	-8.61					0.00		-1519.59	2259
college, ms		15.00	-3.41	-5.16	-8.28	7.38		6.22		1.93				0.00		-1533.77	2259
college, dealer		0.21	0.19	1.04	0.08	0.08		0.03			-0.40			1.38		-1511.37	2259

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
college, dealer.coins		-0.08	0.76	1.08	-0.12	-0.04		0.13				-0.82		3.66		-1524.12	2259
college, grcomp		9.16	-3.02	5.53	0.03	5.65		-3.60					-5.60	0.00		-1502.54	2259
single, ms		15.00	3.68	-9.65	2.16	5.21			-8.01	-3.19				0.00		-1513.55	2259
single, dealer		0.21	0.22	0.96	0.09	0.08			0.02		-0.39			1.35		-1511.41	2259
single, dealer.coins		14.82	-3.00	2.38	-2.88	5.89			-8.61			-9.10		0.00		-1516.32	2259
single, grcomp		15.00	-7.13	4.27	-3.77	2.27			-3.83				-7.26	0.00		-1522.61	2259
ms, dealer		0.09	0.22	1.05	0.07	0.04				0.14	-0.41			1.46		-1510.84	2259
ms, dealer.coins		0.11	0.54	0.85	-0.09	-0.01				-0.02		-0.56		1.48		-1527.27	2259
ms, grcomp		9.02	10.28	-0.91	-9.05	-9.05				0.08			-9.17	1.47		-1529.49	2259
dealer, dealer.coins		0.15	0.61	0.95	0.06	0.06					-0.33	-0.41		1.54		-1510.60	2259
dealer, grcomp		0.20	0.21	1.00	0.08	0.07					-0.38		0.01	1.39		-1511.45	2259
dealer.coins, grcomp		0.10	0.54	0.89	-0.10	-0.03						-0.56	0.02	1.48		-1527.28	2259
female, college, single		123.69	- 101.18	22.45	50.89	220.58	753.41	-13.37	-73.12					0.00		-1490.90	2239
female, college, ms		15.00	2.41	3.81	1.88	-5.16	-7.14	-2.56		-5.34				0.00		-1478.10	2239
female, college, dealer		0.10	0.18	1.26	0.03	0.00	-0.13	0.02			-0.32			2.59		-1496.02	2239
female, college, dealer.coins		15.00	-2.71	5.25	-5.51	0.25	-4.79	-2.85				-7.98		0.00		-1499.91	2239
female, college, grcomp		13.89	-4.44	8.77	0.03	8.87	-0.00	-5.09					-8.84	0.00		-1486.83	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, single, ms		15.00	-3.71	-3.27	3.05	6.02	7.29		-7.82	-3.25				0.00		-1492.03	2239
female, single, dealer		1.05	0.24	12.77	0.90	0.17	-0.75		0.83		-1.23			0.98		-1490.39	2239
female, single, dealer.coins		14.98	-5.07	4.19	-5.04	0.09	-2.51		-4.55			-7.72		0.00		-1493.60	2239
female, single, grcomp		13.87	-11.98	1.67	5.66	14.73	14.35		-7.29				-1.31	0.00		-1491.28	2239
female, ms, dealer		0.06	0.20	1.22	0.05	0.00	-0.11			0.07	-0.35			2.44		-1495.80	2239
female, ms, dealer.coins		2.79	1.05	2.01	-0.17	-1.61	-1.69			-1.56		-0.94		24.32		-1501.80	2239
female, ms, grcomp		14.98	4.64	2.70	-3.25	-9.12	-6.12			-5.99			-3.33	6.08		-1491.62	2239
female, dealer, dealer.coins		0.08	0.47	1.08	0.04	0.01	-0.11				-0.30	-0.29		2.60		-1495.46	2239
female, dealer, grcomp		0.13	0.18	1.35	0.07	0.00	-0.14				-0.35		0.05	2.45		-1495.72	2239
female, dealer.coins, grcomp		14.99	-2.92	4.48	-2.34	-5.48	-9.45					0.37	-5.69	0.91		-1502.84	2239
college, single, ms		15.28	3.55	-10.01	2.38	5.35		-0.11	-7.91	-3.18				0.00		-1511.74	2259
college, single, dealer		15.00	-6.57	3.71	-4.20	0.27		-1.99	-3.90		-4.69			0.00		-1510.86	2259
college, single, dealer.coins		14.90	-4.98	-0.21	-4.73	3.24		3.41	-4.87			-5.15		0.00		-1510.67	2259
college, single, grcomp		9.13	-3.03	5.57	0.10	5.67		-3.62	2.07				-5.56	0.00		-1499.78	2259
college, ms, dealer		15.00	-3.92	-4.50	-9.27	-2.03		6.74		2.82	-4.12			0.00		-1507.90	2259
college, ms, dealer.coins		15.00	6.82	-14.89	-9.05	-2.04		6.47		2.25		-3.67		0.00		-1507.65	2259
college, ms, grcomp		9.63	-3.21	5.72	0.05	5.84		-3.75		-0.16			-5.76	0.00		-1502.17	2259

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
college, dealer, dealer.coins		0.13	0.63	0.97	0.06	0.07		0.03			-0.33	-0.46		1.62		-1510.41	2259
college, dealer, grcomp		11.46	-3.62	7.17	0.03	7.28		-4.26			0.00		-7.23	0.00		-1500.75	2259
college, dealer.coins, grcomp		12.95	0.10	11.66	-4.97	7.35		-6.02				-5.55	-6.97	0.00		-1487.78	2259
single, ms, dealer		15.00	3.23	-9.63	2.27	5.50			-7.76	-3.11	-0.41			0.00		-1509.29	2259
single, ms, dealer.coins		14.93	2.99	-9.77	2.08	5.69			-7.00	-3.22		-0.67		0.00		-1508.98	2259
single, ms, grcomp		13.91	-2.88	-0.32	-3.46	-1.13			-4.85	-2.40			-3.65	0.00		-1502.44	2259
single, dealer, dealer.coins		0.15	0.62	0.91	0.07	0.07			0.02		-0.34	-0.42		1.52		-1510.56	2259
single, dealer, grcomp		0.21	0.22	0.94	0.09	0.08			0.03		-0.39		-0.01	1.34		-1511.41	2259
single, dealer.coins, grcomp		14.49	-3.11	4.58	-3.78	2.27			-7.28			-8.45	-2.33	0.00		-1507.77	2259
ms, dealer, dealer.coins		0.05	0.70	0.97	0.06	0.04				0.13	-0.37	-0.49		1.60		-1509.84	2259
ms, dealer, grcomp		0.09	0.22	1.07	0.07	0.04				0.14	-0.41		0.01	1.46		-1510.83	2259
ms, dealer.coins, grcomp		0.11	0.54	0.87	-0.09	-0.02				-0.01		-0.56	0.01	1.48		-1527.26	2259
dealer, dealer.coins, grcomp		0.15	0.61	0.96	0.06	0.06					-0.33	-0.41	0.01	1.55		-1510.59	2259
female, college, single, ms		19.10	2.92	4.99	2.28	-6.65	-9.01	-3.26	1.70	-7.06				0.00		-1477.37	2239
female, college, single, dealer		15.00	1.20	-10.27	-10.01	-6.15	13.36	8.31	4.22		-1.03			0.00		-1485.39	2239
female, college, single, dealer.coins		14.98	2.82	-11.35	-10.95	-6.68	13.60	9.12	4.62			-1.26		0.00		-1481.91	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, college, single, grcomp		9.56	-14.23	8.90	11.24	7.33	12.48	7.08	-6.26				5.13	0.00		-1465.28	2239
female, college, ms, dealer		18.50	2.60	5.29	0.76	-6.84	-8.94	-2.59		-7.02	1.46			0.00		-1468.45	2239
female, college, ms, dealer.coins		15.00	3.50	1.97	2.10	-4.40	-6.60	-3.86		-4.58		-3.61		0.00		-1474.32	2239
female, college, ms, grcomp		33.70	4.55	6.52	-0.72	-14.48	-13.82	-5.17		-10.20			-4.58	0.00		-1476.05	2239
female, college, dealer, dealer.coins		0.02	0.53	1.12	0.01	0.02	-0.11	0.04			-0.28	-0.39		3.26		-1495.09	2239
female, college, dealer, grcomp		13.89	-4.44	8.77	0.03	8.88	-0.00	-5.09			-0.00		-8.84	0.00		-1486.82	2239
female, college, dealer.coins, grcomp		6.35	6.30	4.74	-1.45	3.95	-0.35	-2.65				-8.50	-3.58	0.00		-1474.82	2239
female, single, ms, dealer		14.54	-11.58	-0.28	6.47	12.12	13.09		-7.77	-1.37	13.55			0.00		-1488.27	2239
female, single, ms, dealer.coins		15.00	-2.06	-8.06	2.34	-2.63	7.55		-4.75	-2.48		7.03		0.00		-1484.56	2239
female, single, ms, grcomp		15.00	-4.81	-0.33	2.05	4.44	14.20		-9.31	-2.61			-4.37	0.00		-1483.94	2239
female, single, dealer, dealer.coins		8.80	-8.00	0.20	3.33	4.93	4.25		-4.16		-8.74	8.24		0.28		-1481.24	2239
female, single, dealer, grcomp		4.16	0.26	5.52	4.03	0.23	-3.90		4.11		-4.26		-0.15	0.70		-1488.98	2239
female, single, dealer.coins, grcomp		12.98	-9.64	3.27	5.57	14.73	14.35		-8.64			0.28	-3.04	0.00		-1487.87	2239
female, ms, dealer, dealer.coins		0.01	0.58	1.06	0.04	0.01	-0.09			0.09	-0.33	-0.40		2.66		-1495.07	2239
female, ms, dealer, grcomp		14.97	4.61	2.72	-3.27	-9.12	-6.12			-5.99	0.06		-3.38	6.20		-1491.35	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, ms, dealer.coins, grcomp		14.96	4.47	2.80	-3.36	-9.17	-6.05			-5.93		0.38	-3.45	6.07		-1491.62	2239
female, dealer, dealer.coins, grcomp		0.08	0.45	1.17	0.06	0.01	-0.12				-0.32	-0.27	0.04	2.63		-1495.21	2239
college, single, ms, dealer		14.74	-3.49	-4.48	-8.63	-1.24		6.97	-0.70	2.54	-3.73			0.00		-1505.17	2259
college, single, ms, dealer.coins		16.66	6.30	-14.17	-10.64	-1.85		6.77	-0.20	2.05		-4.91		0.00		-1506.62	2259
college, single, ms, grcomp		9.23	-3.01	5.57	0.16	5.66		-3.52	2.07	-0.25			-5.50	0.00		-1498.60	2259
college, single, dealer, dealer.coins		0.13	0.65	0.92	0.07	0.08		0.04	0.03		-0.34	-0.47		1.59		-1510.35	2259
college, single, dealer, grcomp		9.71	-3.05	6.21	0.11	6.31		-3.55	2.08		0.00		-6.20	0.00		-1498.01	2259
college, single, dealer.coins, grcomp		5.33	0.14	4.07	-1.20	3.51		-2.24	9.30			-1.77	-3.13	0.00		-1486.53	2259
college, ms, dealer, dealer.coins		15.81	5.17	-13.43	-9.59	-2.08		7.06		2.09	2.56	-6.09		0.00		-1507.28	2259
college, ms, dealer, grcomp		13.84	-13.04	4.32	-6.98	3.14		3.04		8.14	-11.04		2.59	0.01		-1493.92	2259
college, ms, dealer.coins, grcomp		12.74	0.06	11.57	-4.89	7.33		-5.96		0.13		-5.48	-6.96	0.00		-1487.43	2259
college, dealer, dealer.coins, grcomp		6.60	0.16	5.42	-1.78	4.29		-2.92			0.35	-2.71	-3.91	0.00		-1484.79	2259
single, ms, dealer, dealer.coins		13.85	-2.99	-0.33	-4.07	0.10			-3.78	-2.68	7.68	-9.42		0.00		-1496.71	2259
single, ms, dealer, grcomp		14.32	-2.91	-0.33	-3.71	-0.92			-4.91	-2.47	-0.55		-3.22	0.00		-1497.42	2259
single, ms, dealer.coins, grcomp		13.83	-2.95	-0.33	-3.36	-0.72			-4.82	-2.38		-0.48	-3.19	0.00		-1498.02	2259

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
single, dealer, dealer.coins, gr-comp		15.90	-4.14	4.14	-3.65	3.18			-7.78		0.34	-8.24	-3.24	0.00		-1506.20	2259
ms, dealer, dealer.coins, gr-comp		0.05	0.70	0.99	0.07	0.04				0.14	-0.37	-0.50	0.01	1.62		-1509.82	2259
female, college, single, ms, dealer		112.44	36.39	15.16	0.87	-50.26	-54.73	-7.16	50.69	-50.74	6.54			0.00		-1467.32	2239
female, college, single, ms, dealer.coins		23.07	6.22	1.20	3.87	-6.35	-10.26	-6.16	-1.53	-6.75		-6.23		0.00		-1470.11	2239
female, college, single, ms, gr-comp		9.70	-14.10	8.83	11.12	7.55	13.95	7.20	-6.06	-0.39			5.35	0.00		-1459.81	2239
female, college, single, dealer, dealer.coins		6.56	-5.85	0.17	2.80	4.38	3.67	0.07	-3.56		-6.54	6.09		0.29		-1480.51	2239
female, college, single, dealer, grcomp		10.29	-14.69	8.50	10.79	7.65	16.15	7.44	-6.13		0.06		4.82	0.00		-1464.67	2239
female, college, single, dealer.coins, grcomp		10.29	-14.69	8.50	10.79	7.65	16.15	7.44	-6.13			0.06	4.82	0.00		-1464.67	2239
female, college, ms, dealer, dealer.coins		21.73	6.28	1.20	0.82	-6.41	-9.22	-5.80		-6.81	2.31	-8.65		0.00		-1464.57	2239
female, college, ms, dealer, gr-comp		22.01	1.55	6.53	-1.23	-9.03	-9.52	-3.30		-7.21	1.96		-2.03	0.00		-1467.39	2239
female, college, ms, dealer.coins, grcomp		6.89	-6.08	5.23	-7.89	12.89	13.06	3.29		9.10		-11.93	2.52	0.00		-1470.58	2239
female, college, dealer, dealer.coins, grcomp		5.90	6.38	4.45	-1.31	3.80	-0.21	-2.49			0.14	-8.43	-3.42	0.00		-1474.40	2239
female, single, ms, dealer, dealer.coins		9.03	-7.97	0.04	3.45	4.94	4.19		-4.10	-0.46	-8.49	8.30		0.25		-1474.32	2239
female, single, ms, dealer, gr-comp		15.00	-5.92	-0.27	2.49	5.64	13.03		-8.49	-2.82	-0.59		-2.33	0.00		-1481.85	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, single, ms, dealer.coins, grcomp		15.00	-4.81	-0.32	2.02	4.46	14.20		-9.29	-2.60		-0.07	-4.37	0.00		-1483.60	2239
female, single, dealer, dealer.coins, grcomp		9.06	-8.23	0.26	3.21	4.58	4.30		-4.01		-8.80	8.45	-0.27	0.14		-1476.01	2239
female, ms, dealer, dealer.coins, grcomp		14.95	4.48	2.80	-3.35	-9.15	-6.07			-5.94	0.06	0.32	-3.46	6.20		-1491.35	2239
college, single, ms, dealer, dealer.coins		56.86	-8.17	-0.32	-18.58	5.13		-0.00	-22.75	-7.06	0.07	-17.03		0.00		-1492.74	2259
college, single, ms, dealer, grcomp		14.35	-13.22	4.18	-7.16	3.42		3.40	-0.82	7.19	-10.99		3.02	0.02		-1488.15	2259
college, single, ms, dealer.coins, grcomp		11.89	-7.42	-0.29	-6.03	-0.86		7.40	-3.22	5.08		-3.13	-5.92	0.00		-1485.74	2259
college, single, dealer, dealer.coins, grcomp		5.50	0.22	4.31	-1.27	3.69		-2.42	3.05		0.35	-2.20	-3.31	0.00		-1482.72	2259
college, ms, dealer, dealer.coins, grcomp		5.52	0.13	4.27	-1.20	3.73		-2.44		0.05	0.35	-2.12	-3.35	0.00		-1484.69	2259
single, ms, dealer, dealer.coins, grcomp		13.87	-3.00	-0.30	-4.06	0.10			-3.81	-2.71	7.68	-9.42	-0.00	0.00		-1496.70	2259
female, college, single, ms, dealer, dealer.coins		21.56	5.87	1.19	0.66	-6.05	-9.29	-5.92	-1.46	-6.39	2.95	-8.88		0.00		-1460.23	2239
female, college, single, ms, dealer, grcomp		9.83	-14.23	8.85	11.14	7.60	13.95	7.30	-5.99	-0.48	0.22		5.27	0.00		-1454.46	2239
female, college, single, ms, dealer.coins, grcomp		9.89	-14.15	8.84	11.13	7.53	13.95	7.25	-6.01	-0.60		0.24	5.23	0.00		-1454.28	2239
female, college, single, dealer, dealer.coins, grcomp		10.62	-14.89	8.64	10.95	7.56	16.28	7.36	-6.39		0.08	-0.02	4.82	0.00		-1464.56	2239
female, college, ms, dealer, dealer.coins, grcomp		22.44	7.27	1.25	0.87	-7.27	-9.60	-5.40		-7.68	1.68	-9.17	0.19	0.00		-1464.22	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, single, ms, dealer, dealer.coins, grcomp		9.08	-7.99	0.08	3.43	4.77	4.38		-4.09	-0.43	-8.40	8.31	-0.26	0.15		-1469.88	2239
college, single, ms, dealer, dealer.coins, grcomp		5.55	0.25	4.34	-1.30	3.68		-2.42	3.05	-0.07	0.35	-2.23	-3.30	0.00		-1482.58	2259
female, college, single, ms, dealer, dealer.coins, grcomp		9.98	-14.20	8.82	11.22	7.66	13.91	7.32	-5.97	-0.70	0.14	0.21	5.04	0.00		-1453.13	2239
TI only	female	-0.02	0.20	0.90	-0.18	-0.10								3.91	0.00	-1519.34	2239
female	female	0.60	0.07	0.74	-0.13	-0.16	-1.96							0.21	>100	-1506.54	2239
college	female	-0.06	0.20	1.05	-0.19	-0.11		0.03						4.78	0.00	-1519.11	2239
single	female	11.64	-10.35	0.99	3.31	12.63			-4.31					0.00	>100	-1501.33	2239
ms	female	-0.06	0.19	0.99	-0.21	-0.14				0.06				4.66	0.00	-1518.77	2239
dealer	female	0.17	0.22	1.01	0.07	0.07					-0.38			1.63	0.00	-1498.06	2239
dealer.coins	female	-0.01	0.56	0.93	-0.12	-0.02						-0.57		2.82	0.00	-1512.32	2239
grcomp	female	6.99	8.44	-0.87	-7.15	-7.08							-7.19	3.17	0.00	-1515.45	2239
female, college	female	0.45	0.11	0.81	-0.15	-0.14	-2.54	0.05						0.34	>100	-1506.06	2239
female, single	female	14.67	-13.32	1.06	6.51	14.38	9.09		-7.57					0.00	>100	-1501.25	2239
female, ms	female	1.44	0.16	1.36	-0.14	-0.43	-2.01			-0.38				0.06	>100	-1500.20	2239
female, dealer	female	0.43	0.14	0.85	0.01	-0.03	-1.11				-0.26			0.39	95.54	-1491.27	2239
female, dealer.coins	female	14.79	10.99	-0.48	-9.47	-9.36	-6.73					-11.41		0.00	>100	-1494.75	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, grcomp	female	15.00	-2.12	3.68	-1.57	-3.90	-11.04						-4.12	0.00	0.92	-1501.13	2239
college, single	female	30.41	-22.25	8.05	12.42	22.23		-3.50	-20.25					0.00	>100	-1490.91	2239
college, ms	female	1.51	8.27	-7.94	-4.87	-4.92		3.48		3.64				0.29	>100	-1503.48	2239
college, dealer	female	0.21	0.18	1.07	0.09	0.09		0.04			-0.42			1.54	0.00	-1497.96	2239
college, dealer.coins	female	0.01	2.68	-3.29	-0.23	-0.64		1.01				2.50		40.83	>100	-1508.64	2239
college, grcomp	female	7.20	-2.15	4.65	0.25	4.69		-2.75					-4.49	0.00	0.04	-1479.88	2239
single, ms	female	12.25	-7.41	-0.27	3.00	4.96			-5.47	-2.12				0.00	>100	-1485.93	2239
single, dealer	female	14.48	-3.86	2.65	-3.21	4.05			-7.09		-7.62			0.00	2.01	-1495.98	2239
single, dealer.coins	female	11.81	-10.80	0.44	3.67	4.72			-4.44			-0.27		0.02	>100	-1499.47	2239
single, grcomp	female	11.62	-9.73	1.67	4.32	12.63			-5.95				-1.31	0.00	>100	-1491.28	2239
ms, dealer	female	0.01	0.23	1.09	0.05	0.03				0.16	-0.39			2.07	0.00	-1497.24	2239
ms, dealer.coins	female	-0.06	0.58	0.95	-0.13	-0.04				0.04		-0.59		3.15	0.00	-1512.10	2239
ms, grcomp	female	12.45	14.10	-0.89	-12.74	-12.68				0.12			-12.82	4.41	0.00	-1513.09	2239
dealer, dealer.coins	female	0.10	0.64	0.97	0.05	0.06					-0.32	-0.44		1.94	0.00	-1497.12	2239
dealer, grcomp	female	0.17	0.22	1.03	0.07	0.07					-0.38		0.01	1.64	0.00	-1498.05	2239
dealer.coins, grcomp	female	-0.01	0.55	0.95	-0.12	-0.03						-0.57	0.01	2.81	0.00	-1512.31	2239
female, college, single	female	14.99	8.46	-11.41	-13.90	-1.10	-4.96	3.44	1.91					0.01	>100	-1485.36	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, college, ms	female	29.90	4.87	7.05	2.19	-10.85	-13.13	-5.79		-11.02				0.00	0.00	-1477.19	2239
female, college, dealer	female	0.37	0.12	0.94	-0.02	-0.02	-1.34	0.06			-0.25			0.46	>100	-1490.42	2239
female, college, dealer.coins	female	21.51	9.72	-1.27	-6.10	-3.12	-27.56	0.86				-12.78		0.00	>100	-1483.17	2239
female, college, grcomp	female	11.32	-3.58	7.34	0.27	7.37	-0.33	-4.04					-7.15	0.00	0.02	-1478.25	2239
female, single, ms	female	14.94	-8.63	-0.33	2.72	5.29	-3.32		-5.85	-2.85				0.00	>100	-1483.16	2239
female, single, dealer	female	13.28	-4.99	2.86	-4.02	0.75	-4.09		-4.01		-5.30			0.00	1.27	-1482.38	2239
female, single, dealer.coins	female	67.76	-23.44	18.45	-24.92	0.17	-11.60		-19.07			-33.13		0.00	0.00	-1489.97	2239
female, single, grcomp	female	13.87	-11.98	1.67	5.66	15.73	15.35		-7.29				-1.31	0.00	0.00	-1491.28	2239
female, ms, dealer	female	0.39	0.15	0.85	0.01	-0.02	-1.09			0.02	-0.27			0.41	99.71	-1491.22	2239
female, ms, dealer.coins	female	0.89	0.20	0.56	-0.08	-0.09	-2.21			-0.10		-0.34		0.07	>100	-1494.32	2239
female, ms, grcomp	female	14.76	3.89	3.32	-2.91	-8.64	-6.41			-5.84			-3.00	0.53	6.50	-1491.40	2239
female, dealer, dealer.coins	female	0.38	0.30	0.77	-0.01	-0.01	-1.20				-0.22	-0.19		0.42	>100	-1490.40	2239
female, dealer, grcomp	female	0.43	0.14	0.94	0.03	-0.03	-1.17				-0.28		0.04	0.39	>100	-1491.00	2239
female, dealer.coins, grcomp	female	5.46	0.53	2.10	-1.40	1.52	-4.72					-3.88	-1.51	0.00	38.88	-1483.18	2239
college, single, ms	female	12.50	-7.16	-0.13	2.64	4.99		-0.21	-5.21	-2.51				0.00	>100	-1479.76	2239
college, single, dealer	female	14.82	5.85	-14.99	-9.84	-5.97		8.30	4.28		-1.13			0.00	>100	-1485.52	2239
college, single, dealer.coins	female	14.60	2.42	-10.94	-11.09	-6.95		9.31	5.14			-1.51		0.00	>100	-1482.77	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
college, single, grcomp	female	9.44	-12.17	4.41	6.73	4.13		3.93	-3.71				3.24	0.00	>100	-1476.36	2239
college, ms, dealer	female	12.07	0.46	-7.19	-7.24	-1.69		5.81		2.03	-2.41			0.00	>100	-1487.74	2239
college, ms, dealer.coins	female	13.17	8.97	-13.78	-14.84	-3.28		11.16		6.61		-7.76		0.00	>100	-1479.51	2239
college, ms, grcomp	female	9.40	-2.83	6.01	0.47	5.89		-3.47		-0.38			-5.58	0.00	0.04	-1474.55	2239
college, dealer, dealer.coins	female	0.09	0.66	1.00	0.06	0.06		0.03			-0.32	-0.49		2.01	0.00	-1496.94	2239
college, dealer, grcomp	female	16.25	-5.22	10.63	0.28	10.64		-5.86			0.00		-10.43	0.00	0.00	-1478.13	2239
college, dealer.coins, grcomp	female	6.65	11.47	5.29	-1.72	4.24		-2.83				-13.79	-3.86	0.00	0.00	-1477.37	2239
single, ms, dealer	female	15.48	-7.88	-0.29	2.02	5.29			-6.50	-2.84	-0.69			0.00	>100	-1482.13	2239
single, ms, dealer.coins	female	15.48	-7.88	-0.29	2.02	5.29			-6.50	-2.84		-0.69		0.00	>100	-1482.13	2239
single, ms, grcomp	female	12.98	-7.61	-0.33	2.88	5.02			-5.63	-2.30			-0.07	0.00	>100	-1484.77	2239
single, dealer, dealer.coins	female	10.40	-0.58	-9.70	2.04	0.15			-2.15		-10.48	13.61		0.59	>100	-1490.50	2239
single, dealer, grcomp	female	14.92	-6.09	3.54	-1.39	3.18			-7.18		-4.97		-3.25	0.00	8.67	-1482.03	2239
single, dealer.coins, grcomp	female	11.71	-9.65	1.94	4.24	12.63			-6.02			0.20	-1.66	0.00	>100	-1489.34	2239
ms, dealer, dealer.coins	female	-0.04	0.77	1.01	0.04	0.03				0.15	-0.35	-0.56		2.39	0.00	-1496.03	2239
ms, dealer, grcomp	female	0.01	0.23	1.10	0.05	0.03				0.16	-0.39		0.00	2.07	0.00	-1497.23	2239
ms, dealer.coins, grcomp	female	14.20	-9.28	4.93	2.03	5.37				-0.96		0.96	-4.47	0.00	>100	-1508.16	2239
dealer, dealer.coins, grcomp	female	0.10	0.65	0.99	0.06	0.06					-0.32	-0.45	0.01	1.95	0.00	-1497.11	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, college, single, ms	female	29.90	4.87	7.05	2.19	-10.85	-13.13	-5.79	0.00	-11.02				0.00	0.00	-1477.19	2239
female, college, single, dealer	female	22.36	-7.70	6.49	-7.15	0.27	-7.40	-0.05	-7.18		-7.99			0.00	2.40	-1480.24	2239
female, college, single, dealer.coins	female	15.00	1.91	-7.67	-11.87	-4.55	-6.27	9.28	2.76			-4.98		0.00	>100	-1463.01	2239
female, college, single, grcomp	female	12.40	-17.37	9.11	11.41	8.14	10.92	7.89	-6.16				5.45	0.00	0.00	-1464.90	2239
female, college, ms, dealer	female	34.78	8.20	7.02	0.78	-14.12	-16.64	-3.75		-14.44	2.00			0.00	0.00	-1467.39	2239
female, college, ms, dealer.coins	female	15.76	4.32	1.25	2.12	-4.49	-6.71	-4.43		-4.68		-4.43		0.00	0.00	-1474.28	2239
female, college, ms, grcomp	female	9.39	-2.83	6.01	0.47	5.89	-0.09	-3.46		-0.38			-5.59	0.00	0.03	-1473.86	2239
female, college, dealer, dealer.coins	female	651.54	7.74	-1.17	-616.62	-600.16	-72.52	0.82			0.31	-24.56		0.00	>100	-1482.87	2239
female, college, dealer, grcomp	female	16.24	-5.24	10.58	0.21	10.63	-0.09	-5.80			-0.06		-10.40	0.00	0.00	-1477.96	2239
female, college, dealer.coins, grcomp	female	5.82	5.11	12.05	-4.46	4.39	-5.01	4.27				-14.18	3.76	0.00	>100	-1466.97	2239
female, single, ms, dealer	female	13.59	-5.24	3.01	-4.17	0.76	-4.34		-4.19	0.33	-5.56			0.00	1.58	-1480.62	2239
female, single, ms, dealer.coins	female	15.48	-7.88	-0.29	2.02	5.29	-0.00		-6.50	-2.84		-0.69		0.00	>100	-1482.13	2239
female, single, ms, grcomp	female	15.00	-2.00	-0.33	-4.06	-2.29	-0.92		-6.69	-1.93			-5.45	0.00	0.55	-1483.04	2239
female, single, dealer, dealer.coins	female	11.63	-10.82	0.21	4.28	5.81	4.71		-5.09		-11.51	11.05		0.22	2.25	-1480.35	2239
female, single, dealer, grcomp	female	15.15	-6.13	3.57	-1.47	3.26	0.33		-7.28		-5.13		-3.33	0.00	2.50	-1481.91	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, single, dealer.coins, grcomp	female	5.42	0.52	2.06	-1.36	1.53	-4.70		0.12			-3.83	-1.52	0.00	30.46	-1482.06	2239
female, ms, dealer, dealer.coins	female	0.34	0.32	0.78	-0.01	-0.01	-1.19			0.02	-0.23	-0.20		0.45	>100	-1490.35	2239
female, ms, dealer, grcomp	female	0.41	0.14	0.94	0.03	-0.02	-1.15			0.02	-0.28		0.04	0.41	>100	-1490.97	2239
female, ms, dealer.coins, grcomp	female	14.75	6.52	-2.89	-2.50	-4.64	-10.40			-2.26		-6.92	-2.59	0.00	7.03	-1482.04	2239
female, dealer, dealer.coins, grcomp	female	5.44	0.54	2.07	-1.37	1.53	-4.74				0.08	-3.94	-1.52	0.00	37.78	-1482.70	2239
college, single, ms, dealer	female	12.33	-7.28	-0.13	2.66	5.11		-0.21	-4.98	-2.48	0.34			0.00	>100	-1477.15	2239
college, single, ms, dealer.coins	female	13.32	4.76	-11.00	-15.48	-2.71		13.73	0.58	7.87		-8.25		0.00	>100	-1472.31	2239
college, single, ms, grcomp	female	8.98	-12.19	14.26	8.39	6.44		6.10	-4.46	-0.40			4.18	0.00	>100	-1473.59	2239
college, single, dealer, dealer.coins	female	15.25	3.81	-12.68	-11.08	-6.62		9.55	4.82		2.00	-3.22		0.00	>100	-1480.30	2239
college, single, dealer, grcomp	female	10.93	-13.66	7.39	9.69	7.68		6.12	-6.69		-0.02		3.23	0.00	>100	-1466.93	2239
college, single, dealer.coins, grcomp	female	21.36	-24.16	7.38	9.66	7.44		6.07	-6.60			-0.02	3.29	0.00	>100	-1466.87	2239
college, ms, dealer, dealer.coins	female	13.32	4.49	-9.88	-16.68	-2.23		14.03		8.77	0.46	-10.50		0.00	>100	-1477.36	2239
college, ms, dealer, grcomp	female	22.96	-7.56	18.26	0.91	14.73		-8.04		-0.68	0.36		-14.47	0.00	0.00	-1468.64	2239
college, ms, dealer.coins, grcomp	female	6.03	-5.11	4.69	-7.47	8.02		3.52		9.20		-12.15	2.50	0.00	>100	-1464.79	2239
college, dealer, dealer.coins, grcomp	female	5.79	11.58	4.41	-1.22	3.85		-2.54			0.35	-13.68	-3.48	0.00	0.00	-1475.15	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
single, ms, dealer, dealer.coins	female	5.29	2.92	-7.51	1.88	0.11			-2.21	-0.53	-4.75	7.92		0.38	>100	-1479.67	2239
single, ms, dealer, grcomp	female	15.43	-6.41	3.60	-1.60	3.23			-7.28	0.17	-5.29		-3.29	0.00	2.41	-1481.33	2239
single, ms, dealer.coins, grcomp	female	15.48	-7.88	-0.29	2.02	5.29			-6.50	-2.84		-0.69	0.00	0.00	>100	-1482.13	2239
single, dealer, dealer.coins, grcomp	female	14.92	-6.18	3.50	-1.39	3.14			-7.09		-4.89	-0.03	-3.21	0.00	2.63	-1481.94	2239
ms, dealer, dealer.coins, grcomp	female	-0.04	0.77	1.02	0.04	0.03				0.15	-0.36	-0.56	0.01	2.40	0.00	-1496.02	2239
female, college, single, ms, dealer	female	112.37	36.50	15.12	0.81	-50.47	-55.02	-6.67	50.62	-50.72	6.47			0.00	0.00	-1467.32	2239
female, college, single, ms, dealer.coins	female	18.48	1.23	-8.93	-13.95	-4.74	-8.37	9.14	2.42	0.74		-5.07		0.00	>100	-1458.51	2239
female, college, single, ms, grcomp	female	9.86	-14.27	8.78	11.07	7.55	14.03	7.20	-5.98	-0.42			5.43	0.00	0.00	-1459.78	2239
female, college, single, dealer, dealer.coins	female	17.14	2.01	-8.58	-12.44	-3.67	-8.21	8.95	1.87		3.68	-8.62		0.00	>100	-1458.92	2239
female, college, single, dealer, grcomp	female	11.94	1.88	-0.40	-2.94	-2.23	-10.96	5.49	10.60		-5.04		-8.17	0.00	>100	-1462.91	2239
female, college, single, dealer.coins, grcomp	female	10.80	2.11	6.43	-8.83	3.32	-4.90	3.10	1.95			-14.95	2.50	0.00	>100	-1446.61	2239
female, college, ms, dealer, dealer.coins	female	21.63	6.45	1.24	0.82	-6.61	-9.00	-5.80		-6.90	1.82	-8.36		0.00	0.00	-1464.55	2239
female, college, ms, dealer, grcomp	female	34.60	7.66	7.31	0.61	-14.05	-16.58	-3.84		-14.04	2.16		-0.38	0.00	0.00	-1467.33	2239
female, college, ms, dealer.coins, grcomp	female	8.22	3.03	3.82	-6.40	3.33	-5.34	3.12		-0.36		-12.92	2.52	0.00	>100	-1451.50	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, college, dealer, dealer.coins, grcomp	female	3.47	6.07	12.05	-2.18	4.51	-110.61	4.40			0.31	-13.21	3.88	0.00	>100	-1465.89	2239
female, single, ms, dealer, dealer.coins	female	9.03	-7.97	0.04	3.45	4.94	4.19		-4.10	-0.46	-8.49	8.30		0.25	0.00	-1474.32	2239
female, single, ms, dealer, grcomp	female	18.12	-6.49	5.02	-5.91	0.18	-5.91		-5.72	0.34	-6.73		-0.25	0.00	1.39	-1477.59	2239
female, single, ms, dealer.coins, grcomp	female	8.95	0.51	2.09	-1.39	1.54	-8.09		0.11	-0.08		-7.27	-1.53	0.00	21.41	-1481.69	2239
female, single, dealer, dealer.coins, grcomp	female	11.45	-10.59	0.28	6.26	7.54	6.74		-7.04		-11.11	10.79	-0.30	0.10	2.32	-1474.36	2239
female, ms, dealer, dealer.coins, grcomp	female	39.48	6.42	-3.50	-3.04	-4.46	-35.30			-1.57	0.12	-9.43	-3.17	0.00	6.71	-1481.02	2239
college, single, ms, dealer, dealer.coins	female	14.24	7.74	-13.62	-19.31	-3.19		16.25	0.48	10.61	0.46	-12.20		0.00	>100	-1470.32	2239
college, single, ms, dealer, grcomp	female	12.69	-14.88	8.00	10.29	7.73		7.19	-6.00	-1.82	0.47		4.10	0.00	>100	-1451.83	2239
college, single, ms, dealer.coins, grcomp	female	165.67	-164.68	142.55	-276.42	152.44		134.57	0.51	296.25		-430.20	133.52	0.00	>100	-1456.96	2239
college, single, dealer, dealer.coins, grcomp	female	21.35	-24.16	7.39	9.67	7.44		6.07	-6.60		-0.01	-0.02	3.29	0.00	>100	-1466.87	2239
college, ms, dealer, dealer.coins, grcomp	female	81.51	-80.52	72.42	-96.28	28.82		24.63		121.96	-66.92	-79.07	23.51	0.00	>100	-1463.74	2239
single, ms, dealer, dealer.coins, grcomp	female	9.31	-2.82	-5.64	1.97	-0.06			-2.35	-0.53	-8.45	11.80	-0.40	0.17	>100	-1475.61	2239
female, college, single, ms, dealer, dealer.coins	female	19.01	1.48	-9.16	-15.19	-4.56	-8.49	9.66	2.27	1.73	0.56	-6.72		0.00	>100	-1453.88	2239
female, college, single, ms, dealer, grcomp	female	12.87	-14.92	8.01	10.29	7.55	2.21	7.03	-6.11	-1.87	0.56		3.82	0.00	>100	-1450.11	2239

r-variables included (in addition to treatment indicators)	σ-variables included	r												σ			
		constant	coins	ungraded	skewLO	skewHI	female	college	single	ms	dealer	dcoins	grcomp	constant	female	LL	N
Full Model (original estimates)	female	0.95	-0.16	3.75	0.76	0.14	-1.26	0.04	0.84	0.03	-0.98	0.39	-0.12	0.41	12.54	-1486.29	2259
female, college, single, ms, dealer.coins, grcomp	female	10.32	2.42	6.18	-8.96	3.93	-5.07	3.46	1.10	0.50		-15.27	2.54	0.00	$>$100	-1439.93	2239
female, college, single, dealer, dealer.coins, grcomp	female	11.44	1.55	6.93	-9.70	3.88	-4.75	3.47	4.87		0.36	-15.39	2.53	0.00	$>$100	-1440.04	2239
female, college, ms, dealer, dealer.coins, grcomp	female	11.42	2.62	6.66	-9.43	3.83	-49.29	3.45		-0.27	0.36	-16.17	2.55	0.00	$>$100	-1447.77	2239
female, single, ms, dealer, dealer.coins, grcomp	female	9.08	-7.99	0.08	3.43	4.77	4.38		-4.09	-0.43	-8.40	8.31	-0.26	0.15	0.00	-1469.88	2239
college, single, ms, dealer, dealer.coins, grcomp	female	12.65	-14.82	8.02	10.30	7.65		7.13	-6.02	-1.85	0.50	0.04	4.09	0.00	$>$100	-1450.49	2239
Full Model (best estimates)		10.18	2.31	6.04	-8.98	3.94	-5.11	3.56	1.11	0.52	0.37	-15.52	2.71	0.00	$>$100	-1438.88	2239

Each row shows the parameter vector with the highest log-likelihood amongst those found using the methods described in Section 1.A4. Cells with parameters determining r shaded gray if parameter is larger than 3 in absolute value.

Table 1.A2: Results of numerical maximization for all possible specifications using the set of covariates used by HLT

1.A5 Additional issues

The econometric analysis of HLT contains a few additional issues not mentioned in the main body of this Comment. They are listed below.

Prospect Theory

In addition to the CRRA expected utility + Fechner error model that is the subject of the main text of this Comment HLT also estimate a classical (i.e. non-cumulative) prospect theory model on their data. HLT report (emphasis in the original):

> *The prospect theory specification mitigates the quantitative effects of the treatments, but does not change the main qualitative conclusions.* For parsimony, we only summarize the results here. The probability weighting parameter γ is estimated to be 0.83, which is significantly less than 1 and consistent with previous estimates from laboratory experiments. The effect of the graded coins treatment on risk aversion is again statistically insignificant, after controlling for the same characteristics used in Table II. The CRRA parameter α increases by 0.65 with graded coins, but has p-value of 0.86. On the other hand, α increases by 0.59 with ungraded coins, and has a standard error of only 0.18 and a p-value of 0.001. The 95% confidence interval on this effect from background risk is between +0.24 and +0.94, so the treatment effect magnitude is much smaller than under EUT, but we find qualitatively similar insights.

None of these results are to be found in the log file in the paper's supporting material (see Listing 1.2 for the relevant excerpt. Note that the log-likelihood of this model, a model in which HLT's baseline specification is nested, is lower than that of HLT's baseline specification) nor have I been able to reproduce them. Indeed, attempts to try to reproduce them yield slightly higher likelihoods than the baseline model but point estimates for γ which are *larger* than one (albeit not statistically significantly differently so), for a probability weighting function that is s-shaped rather than inverse s-shaped.

```
                                                  Number of obs   =       2239
                                                  Wald chi2(6)    =          .
Log pseudolikelihood = -1510.9798                 Prob > chi2     =          .

                                   (Std. Err. adjusted for 112 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
crra         |
       coins |   .9284791   .1856117     5.00   0.000     .5646869    1.292271
    ungraded |   1.203079   .7054548     1.71   0.088    -.1795867    2.585745
      skewLO |   9.316468          .        .       .            .           .
      skewHI |   .2497722          .        .       .            .           .
      female |   .3728553          .        .       .            .           .
     college |    .212427   .3683002     0.58   0.564    -.5094282    .9342822
      single |  -.1756582   .1483901    -1.18   0.237    -.4664976    .1151811
          ms |   1.302932   4.758075     0.27   0.784    -8.022724    10.62859
      dealer |   -3.63777   .3345559   -10.87   0.000    -4.293488   -2.982053
     d_coins |  -.7240507          .        .       .            .           .
      grcomp |  -.2091023          .        .       .            .           .
       _cons |   2.366453   4.944522     0.48   0.632    -7.324633    12.05754
-------------+----------------------------------------------------------------
gamma        |
       _cons |   1.251066   .4030218     3.10   0.002     .4611582    2.040975
-------------+----------------------------------------------------------------
noise        |
      female |   6.967949   29.98527     0.23   0.816    -51.80211      65.738
       _cons |   46.31619          .        .       .            .           .
------------------------------------------------------------------------------
```

Listing 1.2: Estimation results for the Prospect Theory model. Excerpt from `coins.log`, lines 1961 – 1990

Numerical instability

The do-files in Harrison et al.'s supplementary material do not allow for exact replication for many of the results given in the paper (even the numbers in the original log-file that is part of the paper's supplementary material often do not exactly match those reported in the paper) though results are often similar. Strangely, in some cases this seems to be caused by a numerical instability that is due to some interaction between the weak identification of some parameters and the sorting of the data set.

In `coins_ml.do`, line 16 the data set is sorted on a variable that codes for the skewness frame and the index for the decisions in the multiple price list:

```
sort frame decision
```

This combination of variables does not uniquely determine the sorting of the data set since multiple participants got to see the same skewness frame. Since Stata's sorting algorithm is a random sort, all ties are resolved randomly. And since the state of the random number generator which governs this sort will differ in successive runs so will the sorting of the data set. How *exactly* the data is sorted then influences the precise numerical value the log-likelihood function, which is just a sum of individual likelihoods, returns since in floating point arithmetic the results of addition may be different depending on the order of the summands, as the following Stata code example demonstrates:

```
. di %25.0g 0.3+0.6+0.1
    .99999999999999989
. di %25.0g 0.3+0.1+0.6
    1
```

These differences in the calculated log-likelihood are often minuscule but because the log-likelihood is so flat around the estimated coefficient for the *Ungraded coins as Final Outcomes* treatment dummy these small differences can be mirrored in some fairly sizable differences in the *Ungraded coins as Final Outcomes* coefficient. The instability is likely what is behind the difference between the parameter value reported in the paper and the value in the log-file included in the supplementary material.

Making the sort unique by replacing the line above with

```
sort frame decision id
```

eliminates the difference between successive runs as does fixing the seed that governs the sort via `set sortseed` or executing the script on a fresh Stata session (since the random number generator is always initialized with the same `sortseed` of 1001). It does not, however, eliminate the root cause of the issue, the weak identification of one of the parameters to be estimated.

A miscoded observation

The coding of the indicator variable which codes for whether or not a subject is "`single`" leads to an observation ($id = 60$) whose marital status ("`marital`") is missing in the raw data set to be coded as not single. From `coins_data.do`, lines 227–228:

```
gen single=0
replace single=1 if marital==1
```

This miscoding does not much matter if the set of variables included in the model is sufficiently small. HLT's original model estimates, however, are sensitive to the coding of this single observation. Changing the variable to missing (which removes it from the subsequent analysis) or, alternatively, changing it to 1 leads to the pathology described in Section 1.2.2: The constant is estimated to be more than 5, the treatment effect of using coins instead of monetary prizes is sharply negative and the treatment effect of using ungraded instead of graded coins grows even larger than in the original specification. A closer look reveals that this is driven almost entirely by a single subject ($ID = 64$) who chooses the comparatively risky option B throughout, whose CRRA parameter is therefore unbounded below and who gets separated after the recoding. Reducing the set of covariates to the ones in the reduced specification remove this instability.

2 The Standard Portfolio Choice Problem in Germany

based on work with Steffen Huck & Georg Weizsäcker

We report on an artefactual field experiment that examines investment behavior in a representative sample of the German population. The experiment uses households from the Socio-Economic Panel's "Innovation Sample" (SOEP-IS) as respondents. They act as investors who face a standard portfolio choice problem, allocating a fixed budget between a safe and a risky asset. No other investments are possible and the investment horizon is fixed. Despite its drastic simplification, the standard portfolio choice problem is widely viewed as capturing one of the main tradeoff in financial decision making. We regard its relevance as an empirical question and examine both its internal consistency and external validity for the German general population. Regarding external validity, behavior in our artefactual investment task is robustly correlated with actual stock market participation, even after controlling for many of the correlates of participation that the existing literature has identified. The average stock market participation rate is 18% in our representative sample of households; and a one-standard-deviation increase in the experiment's investment in the risky asset is associated with an increase in stock market participation by 6 percentage points. Regarding internal consistency, we find that investments in the risky asset are correlated with measures of beliefs about the asset's return, lending further credibility to the story that the standard portfolio choice model sets out to tell. However, the data also shows how severely respondents' cognitive limitations and financial skills affect decisions. We exogenously vary the returns of the risky asset across treatment groups, by paying some groups a fixed percentage over and above the stock market return and some groups a fixed percentage below the stock market return, and find that only a subsample of relatively well-educated respondents reacts to such changes in incentives. For all other respondents, the opportunity to earn additional money is lost.

Alongside the artefactual field experiment, we also present a laboratory study in which we use the same protocol on a convenience sample of university students. The results are largely congruent between the two settings, with one notable difference: unlike the general population, university students *do* react to the variation in incentives.[1]

The above evidence points at an important role of task complexity for financial decision making. It is plausible that university students understand the incentive structure better and that this induces them, but not the typical German household, to react to the incentive change. That is, the perceived complexity of the incentive change may be higher for some people than for others. We also examine another channel. The two assets in the standard portfolio choice problem differ in nature, one being characterized by just a single number, the other by a (subjective) probability distribution. An investor may find it easier to appreciate a shift in the single number than in the probability distribution.

We test this new hypothesis in an additional laboratory experiment where economically equivalent incentive shifts come in two guises—once as a shift in the return of the risky asset and once as a shift in the safe return. The experimental design ensures that both incentive variations are equally easy to understand[2] and each participant faces both kinds of manipulations. The experimental results confirm that the reaction to changes in the safe asset is indeed significantly stronger than the reaction to changes in the risky asset. This pattern has not yet been observed in the literature, to our knowledge, and cannot be explained by standard theories of decision-making under uncertainty.[3] It has, potentially, important consequences for the optimal design of tax incentives for investments

[1] Notice, however, that the students, just as the SOEP participants, exhibit too mild a change in beliefs in response to incentive changes.

[2] The experiments involve incentive shifts for both assets, presented in the same format. A controlled variation of the shift sizes and a simultaneous variation of an illiquid asset generates the isomorphy within pairs of incentive shifts.

[3] One possible way of rationalizing the pattern is to posit that the manipulation of the assets affects the perceived source of uncertainty (in the sense of Fox and Tversky (1995), and Abdellaoui, Baillon, Placido, and Wakker (2010)).

and other regulatory measures.

Relation to existing literature. Our experimental design builds on the sizable literatures on stock market participation, belief elicitation and experiments on choice under uncertainty. Our results are mostly, but not in all cases, consistent with these literatures and we emphasize some of the relevant comparisons.

The observation that stock market participation is puzzlingly low is widely credited to Haliassos and Bertaut (1995) who find that not only do relatively few members of the middle class invest in stocks, but even amongst the rich, where classical rationales for non-participation are unlikely to hold, participation is far from universal. Germany is a strong case for this puzzle, with its low percentage of stockholders. “Behavioral” explanations of the puzzle are common in the literature[4] and observational or experimental findings on financial literacy and subjective expectations abound (for survey evidence on financial literacy and its correlates in the German population, see Bucher-Koenen & Lusardi, 2011).

A growing literature measures the subjective beliefs of the general public about stock returns. The earliest survey questions asked for a measure of central tendency only (Vissing-Jorgensen, 2004). Questions to elicit entire distributions have more recently been added to many surveys.[5] These questions ask for statements about the probabilities of the market returns lying above given thresholds.[6] The broad picture emerging from this literature is that expectations are extremely heterogeneous, often lie far away from actual returns (Hurd et al., 2011)[7] and show

[4] Frequently mentioned explanations are education, cognitive skills (Grinblatt, Keloharju, & Linnainmaa, 2011) and financial literacy (van Rooij, Lusardi, & Alessie, 2007), transaction cost and availability of information, and ambiguity aversion (Dimmock, Kouwenberg, Mitchell, & Peijnenburg, 2013).

[5] See the Survey of Economic Expectations (Dominitz & Manski, 2011), the Michigan Survey of Consumers (Dominitz & Manski, 2011), the American Life Panel (Hurd & Rohwedder, 2012), the French ‘Mode de vie des Français’ panel (Arrondel, Calvo-Pardo, & Tas, 2012) and the Dutch CentER panel (Hurd, van Rooij, & Winter, 2011).

[6] In the Health and Retirement Survey respondents are asked for the “chance that mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth more than they are today” and the “chance they will have grown by 10 percent or more” (Dominitz & Manski, 2007). Assuming no measurement error these two questions yield two points on the CDF and, if one is willing to make distributional assumptions, allow fitting an entire distribution for every individual.

[7] For example, Kézdi and Willis (2009) find that in 2002 the average subjective probability of a stock market gain was just 49% compared to a historical frequency of 73%. Dominitz and Manski (2011) report that from 2002 to 2004, the average subjective probability of a gain was

positive predictive power for stock market investments.

One drawback of these methods is that responses are often internally inconsistent (Binswanger & Salm, 2013).[8] Instead of asking for probabilities of a return lying above a threshold, we use a histogram elicitation method pioneered by Delavande and Rohwedder (2008) in which respondents are asked to distribute a fixed number of items that jointly represent a probability mass of 1 into a number of bins. The method allows using all available data instead of focusing on consistent sets of responses. The method also has the advantage of being easy to understand; it has been successfully used even with respondents with little formal education and low numerical and statistical skills (Delavande, Giné, & McKenzie, 2011).[9]

In contrast to previous findings, the respondent in our sample report beliefs that accurately capture the historical market return distribution, at least in the aggregate. This is detailed in Appendix 2.A1. A further notable difference is that while experimental investments have high external validity in our sample, the elicited beliefs have much less predictive power for stock market participation. This may in part be due to our comparatively small sample size as well as to the different parts of the sample which enter into the econometric analysis (previous studies often discard the sizeable numbers of respondents who report internally contradictory beliefs). But there is further evidence suggestive of a systematic difference between the German sample and others: the subjective probability of the relevant stock market index making a gain varies significantly less between stockholders and non-stockholders in our data than it does in the other studies.[10]

While there is a large literature on how people make risky choices[11] and on the

46.4%.

[8] In the Health and Retirement Survey 41% of respondents give the same answer to both the question about the likelihood of a positive return and the question about a return above 10%, and a further 15 % violate monotonicity outright.

[9] We additionally ask respondents for a simple numerical expectation, which yields very similar results in most parts of the analysis.

[10] In each of Hurd et al. (2011), Dominitz and Manski (2011) and Arrondel et al. (2012), the stockholders assign about ten percentage points more probability mass to the event that the relevant index makes a gain. In our data, this probability differs between stockholder and non-stockholders only by 2.3 percentage points.

[11] For evidence on choice patterns in representative samples, see, e.g. Andersen, Harrison, Lau, and Rutström (2008), Rabin and Weizsäcker (2009), von Gaudecker, van Soest, and Wengström (2011), Huck and Müller (2012) or Choi, Kariv, Müller, and Silverman (2014).

relevant correlates[12], there are no existing studies that we know of that examine whether risky choices in simple lab-style portfolio problems help to predict stock holdings. But while our finding of a strong correlation between an experimental investment and real-life stock market participation is new, the idea is not. In the working paper version of Dohmen et al. (2011) the authors report on an investment experiment that was also done in a German household survey but is simpler than ours. Dohmen et al. make the important observation that domain-specific risk attitudes are better predictors of real-world behavior than general risk attitudes. This is consistent with our finding that a choice framed in the context of financial markets is a better predictor for real-life stock holdings than, for example, the respondents' general risk tolerance.

There is also a growing literature on how the complexity of the choice environment can produce suboptimal choices and muted reactions to changes in incentives. Wilcox (1993) and Huck and Weizsäcker (1999) present laboratory experiments showing that complexity of simple lotteries affects lottery choices. Chetty, Looney, and Kroft (2009) show that consumers react to the inclusion of sales taxes on price tags even when the after-tax price of goods does not change and react more weakly to changes in taxes that are applied at the register instead of being posted on the price tag. Abeler and Jäger (2015) find much the same thing in a laboratory real-effort task in which earnings are taxed either according to a straightforward schedule or a more complex schedule, which is described by 30 rules. Though both schedules yield the same optimal work effort in theory, subjects who face the complex schedule are further away from the optimal solution. Moreover, and similar to our findings, participants with comparatively low cognitive abilities react less strongly to the imposition of new tax rules under the complex schedule.[13]

The remainder of the paper is organized as follows. In Section 2.1 we describe

[12] For example, Guiso, Sapienza, and Zingales (2008) show with Dutch household panel data how general trust correlates with stock holdings.

[13] We note that given the lack of response to stark variations in incentives that we observe in our study, it is perhaps not surprising that, elsewhere, investors are found to react to extraneous information such as advertisements for standard financial assets (like individual stocks) or photos of financial advisors (Bertrand, Karlan, Mullainathan, Shafir, & Zinman, 2010). This is also consistent with the findings of Binswanger and Salm (2013) who argue that large subsamples of the population may not think probabilistically about stock market returns at all.

the experimental design and procedures for both the household panel and the laboratory. In Section 2.2 we focus on the experimental data and study the relation between beliefs about returns and investments in the experiment. In Section 2.3 we turn to the validity questions that relate the experimental data to socioeconomic data from the household panel, and in Section 2.4 we examine the treatment effects. Section 2.5 presents the additional experiment comparing the return manipulation between safe and risky assets, and Section 2.6 concludes.

2.1 Experimental Design and Procedures

2.1.1 Survey module

Our experimental module was part of the 2012 wave of the German Socioeconomic Panel's Innovation Sample (SOEP-IS). The SOEP is a nationally representative sample of the German population and the SOEP-IS is its sister survey which is used to trial new questions and modules (see Richter & Schupp, 2012, for details). Its sampling of households follows the same procedure as the SOEP does and renders the SOEP-IS representative of the German population. The module was presented to 1146 respondents in 700 households, all of which were added to the SOEP-IS sample in 2012. All households completed the SOEP baseline questionnaire on the same day as our experimental module. Trained interviewers collected responses via computer-aided personal interviewing (CAPI) at the respondents' homes. In the data analysis, we will only use the responses from the "head of household", whom we take to be the household member who responds to the household questionnaire in addition to the personal questionnaire that every household member answers.

Our module contains a regular survey component that we use to elicit several aspects of respondents' asset portfolio (liquid assets, debt, retirement savings) as well as financial literacy and attitudes towards savings and risk. The core component of the module is the interactive experiment modeled on the standard portfolio choice problem that we describe in the following.[14][15]

[14] To minimize interviewer influence, the CAPI-notebooks are placed in front of the respondents and they themselves get to enter their responses. Interviewers are instructed to intervene only if respondents show visible difficulties with the task or explicitly ask for help.

[15] A complete set of instructions is available in the Supplementary Material.

The first screen of our experiment shows respondents a summary description of the investment decision. They are asked to imagine owning €50,000 that they will invest for the duration of one year. The two available assets are a safe asset that pays 4% and is framed as a German government bond, and a risky asset, referred to as the "fund". The fund is based on the DAX, Germany's prime blue chip stock market index. Respondents receive a one-sentence description of the DAX and learn that, depending on the treatment, the fund pays a return equal to a DAX return drawn from the historical distribution plus a percentage point shifter. There are five treatments that differ in the value of the shifter, with possible values in the set $\{-10, -5, 0, 5, 10\}$. Respondents are randomly allocated to treatments. If their shifter value is 0, then the shifter is not mentioned (for simplicity). Otherwise the first screen indicates the absolute size of the shifter but not its sign. For example, a respondent would learn that the fund pays either 5 percentage points less than the DAX or 5 percentage points more than the DAX and that she will subsequently learn which of the two values applies. The respondents also learn that they will be paid in cash on a smaller scale at the end of the survey.

On the second screen, respondents receive more detailed explanations about the determination of payments including (in bold letters) the information of the shifter's sign that "the computer has determined through a random draw". We use this two-step revelation of the shifter's random draw in order to maximize the respondent's appreciation that the shifter is random with zero mean, carrying no information about the underlying DAX return. Since each respondent is only confronted with one realized shifter value in their choice problem, showing the mirrored value makes it salient that the shifter carries no information. The procedure also ensures that the instructions of the laboratory replication are identical despite the fact that only two shifter values are possible there (see Section 2.1.2 below).

The text on the second screen also gives some numerical examples and specifies that the fund's return depends on a draw from historical DAX returns from 1951 to 2010 and that actual payments are scaled down by a factor of 2000.[16]

[16] For all years since the DAX's origination in 1988 we use the actual yearly returns on the index. For all previous years we make use of the yearly return series from Stehle, Huber, and Maier (1996) and Stehle, Wulff, and Richter (1999), who impute the index going back all the way to 1948. All returns are nominal. In contrast to e.g. the S&P 500 the DAX is a performance

Upon reading these short instructions the respondents make their investment decision on the third screen. Respondents who invest their entire endowment in the riskless asset would receive a certain payment of €26. Investing the entirety in the risky asset could yield a payment anywhere from €11.52 to €56.52 depending on the treatment and the randomly drawn year. No information on historical returns is made available to the respondents during the experiment. Under the assumptions of rational expectations, EU-CRRA and usual degrees of risk aversion, one can generate the approximate prediction that in treatments with non-negative shifters, all respondents with degree of relative risk aversion below 3 should invest their entire endowment in the risky asset; those with a shifter of -10 should invest very little whereas those with -5 should invest intermediate amounts.[17]

On the fourth screen we elicit respondents' beliefs about the return of the fund, using the histogram elicitation method pioneered by Delavande and Rohwedder (2008) and refined by Delavande et al. (2011) and Rothschild (2012).[18] A screenshot of the interface can be found in Appendix 2.A4. Respondents have to place 20 "bricks", each representing a probability mass of 5%, into seven bins of possible percentage returns. The set of available bins is {(-90%,-60%),(-60%,-30%),(-

index, which means that dividend payments are included in the return calculations.

[17] These statements hold in a classic two-period two-asset portfolio choice model with log-normal asset returns and CRRA utility over wealth in the second period (i.e. a simplified version of Merton (1969) and Samuelson (1969); see also Campbell and Viceira (2002)). In this model the optimal stock investment share α can be approximated by

$$\alpha = \frac{\mu_r - r_f + \sigma_r^2/2}{\rho \cdot \sigma_r^2},$$

where μ_r is the expected log return, σ_r^2 is the variance of returns, r_f is the natural logarithm of the risk-free rate and ρ is the coefficient of relative risk aversion. Over the payoff-relevant period 1951-2010 the log-normality assumption was approximately correct for year-on-year returns on the DAX (Shapiro test p-value: 0.6), the mean log-return was 0.11 and the variance of returns was 0.1. The riskless asset in the experiment paid 4%. The predictions made in the main text readily result under rational expectations. For respondents with log-utility ($\rho \approx 1$) the optimal stock investment share in Treatment 0 is 1, in Treatment -5 it is 0.74 and in Treatment -10 it is 0.22. Under the same assumptions positive shifters have no effect on stock investment, which remains at the corner solution. However, given that stock investments observed in reality are often much lower than those predicted by the model and that most of the finance literature estimates risk aversion to be substantially higher we decided to also include positive shifters.

[18] For an overview of studies which have used this or similar methods see Goldstein and Rothschild (2014) and references therein.

30%,0%),(0%,30%), (30%,60%),(60%,90%),(90%,120%)}. The bins are, hence, wide enough to allow responses over the entire historical support of DAX returns[19] and, more generally, allow for a large set of possible subjective beliefs. In addition, on the fifth screen, respondents enter the "average return [they] expect for the fund". For both the histogram elicitation of beliefs and for the stated beliefs, it is straightforward to formulate the rational prediction of treatment differences: no matter the distribution of beliefs in the population, the shifter should move beliefs one-to-one. For example, between the -10 shifter and the +10 shifter treatments reported beliefs should differ by 20 percentage points.

Like all previous surveys on beliefs about stock market returns we decided not to incentivize either of these belief measures. Properly incentivizing subjects would have required a payment mechanism whose explanation would have strained the attention span of our respondents (see Allen, 1987, for an example of such a mechanism) and taken up valuable survey time for very little gain.[20]

On the sixth and seventh screens, respondents report how confident they are of their belief statements, on a scale from 0 ("not at all") to 10 ("very sure"), and answer a few understanding questions. The eighth screen elicits the respondents' beliefs about next year's DAX return using the same histogram interface that was used before. Finally, on the ninth and last screen of the experimental module respondents were told which of the years between 1951 and 2010 had been drawn and received a detailed calculation for their payment. Respondents were paid in cash, with amounts rounded up to the nearest euro, at the end of the entire survey interview. On average respondents received €27.16 (min: €17, s.d.: €3.43, max: €48).

Before respondents are presented with the experimental module and its instructions, they have a choice whether or not to participate. The participation rate is

[19] The lowest return on the DAX in the payoff-relevant period was -43.9% in 2002. The highest return was 116.1% in 1951. The lowest bin was included for reasons of rough symmetry and to keep subjects from anchoring their reports on the lowest possible return displayed in the interface.

[20] Both Armantier and Treich (2013) and Trautmann and van de Kuilen (2015) show that the wrong scoring rule can induce bias in the responses. In contrast, not incentivizing the elicitation of beliefs does not yield biased answers in these studies but merely noisier answers. A further concern with incentives is the introduction of possible motives for attempted hedging between tasks (see e.g. Karni & Safra, 1995).

	Dependent variable: Participation in the Experiment
Female	−0.001 (0.030)
Born in the GDR	0.028 (0.038)
Abitur	0.043 (0.058)
University Degree	−0.001 (0.070)
Household Size	−0.018 (0.019)
Number of Children in Household	0.019 (0.034)
Employed	0.017 (0.038)
Financially Literate	0.028 (0.030)
Interest: < 250 Euros	−0.028 (0.035)
Interest: 250 - 1.000 Euros	0.027 (0.049)
Interest: 1.000 - 2.500 Euros	0.096 (0.093)
Interest: > 2.500 Euros	0.120 (0.240)
Interest: refused to answer	−0.076 (0.087)
Stock Market Participant	0.025 (0.046)
Risk Tolerance: Low	0.029 (0.033)
Risk Tolerance: High	0.027 (0.041)
Age bracket 31-40	0.032 (0.077)
Age bracket 41-50	−0.083 (0.059)
Age bracket 51-60	−0.084 (0.057)
Age bracket 61-70	−0.064 (0.060)
Age bracket > 70	−0.200*** (0.059)
N	692

*p < .1; **p < .05; ***p < .01
Standard errors are bootstrapped with 1000 replicates

Table 2.1: Selection into the experiment: Probit marginal effects

80%. Those who decline primarily cite old age and problems with using computers but also a lack of interest in financial matters or ethical or religious reservations against any sort of financial "gambling". The probit regression shown in Table 2.1 mirrors these answers from the open-ended question about the reasons for non-participation. The most potent predictor, indeed the only predictor, of selection into the experiment is age. Respondents over the age of 40 are somewhat less likely to participate and respondents above the age of 70 are significantly less likely to participate though almost two thirds in this age group still participate. All other observable characteristics play no role in the selection into the experiment. A Wald-test for the joint significance of all variables other than the age brackets cannot reject the null of no effect ($\chi^2(18) = 19.41$, $p = 0.37$).

2.1.2 Laboratory Experiment

Upon completion of the field data collection in the SOEP-IS, we used the identical experimental module for a set of 198 university students in the WZB-TU Berlin decision laboratory. Recruitment into the laboratory sample followed standard procedures.[21] The instructions and sequence of informational displays on the computer screens in the laboratory were as close to the CAPI environment as we could produce them, so that the potential practical difficulties with the format would affect both populations. The experimental participants' payments were also scaled by the same factor as payments to SOEP participants. The only relevant difference in experimental design and procedures are that (i) the experimental participants do not have to fill out the long SOEP questionnaire, and (ii) we conducted only two treatments with return shifters -10 and 10, in the laboratory, focusing on the strongest treatment difference in incentives. Since the SOEP respondents who happened to be in either of these two treatments were only informed about the existence of these two treatments, we could leave the instructions entirely unchanged between survey and lab environments.

2.2 Experimental Data

2.2.1 Beliefs and Investments

We start with a summary description of investments and elicited beliefs. We will call the share of wealth a respondent invests in the fund "equity share" hereafter. In both samples the distributions of equity shares have relatively wide supports and few people invest all or nothing. Summing over all treatments, the means (and standard deviations in parentheses) of the equity share are 0.37 (0.25) in the SOEP sample and 0.46 (0.31) in the laboratory sample. The proportions of respondents investing all, exactly half, or nothing in the risky asset are 0.03, 0.2 and 0.18 in the SOEP sample and 0.12, 0.05 and 0.09 in the laboratory sample.

A description of the beliefs about the fund's return is more involved, since each belief report consists of an entire histogram. A clear difference between the SOEP

[21] The decision laboratory uses ORSEE (Greiner, 2015).

	Equity Share		Imputed Expectation of Belief		Imputed S.D. of Belief		Stated Expectation of Belief		
	Mean	S.D	Mean	S.D	Mean	S.D	Mean	S.D	N
Overall	0.37	(0.25)	12.53	(20.59)	23.96	(16.54)	8.27	(17.84)	562
Age Bracket									
<30	0.41	(0.27)	12.16	(16.06)	30.25	(16.07)	8.74	(16.64)	82
31-40	0.39	(0.22)	13.85	(15.73)	25.60	(17.13)	12.02	(16.54)	76
41-50	0.40	(0.23)	12.57	(24.70)	26.36	(16.75)	7.12	(18.65)	107
51-60	0.37	(0.26)	13.24	(21.86)	22.72	(16.46)	8.43	(19.41)	107
61-70	0.34	(0.26)	10.02	(19.63)	20.46	(15.88)	6.22	(17.27)	111
>70	0.32	(0.28)	14.13	(22.49)	19.19	(14.77)	8.36	(17.63)	79
Gender									
female	0.35	(0.24)	9.72	(22.29)	25.60	(17.20)	7.86	(21.59)	271
male	0.39	(0.26)	15.14	(18.52)	22.43	(15.78)	8.65	(13.46)	291
Born in									
West Germany	0.37	(0.26)	12.11	(20.97)	23.34	(15.60)	7.40	(17.38)	379
East Germany	0.34	(0.23)	12.87	(21.96)	22.47	(17.46)	7.75	(17.69)	116
abroad	0.42	(0.28)	14.95	(15.44)	29.74	(19.10)	14.66	(17.35)	54
Abitur									
yes	0.37	(0.28)	10.74	(19.51)	26.70	(14.83)	6.40	(13.47)	122
no	0.37	(0.25)	13.02	(20.87)	23.20	(16.93)	8.78	(18.85)	440
University Education									
yes	0.35	(0.28)	11.54	(21.78)	26.95	(15.40)	5.55	(16.46)	72
no	0.37	(0.25)	12.67	(20.42)	23.52	(16.67)	8.67	(18.01)	490
Employed									
yes	0.39	(0.25)	13.64	(20.70)	24.38	(16.13)	8.98	(16.13)	297
no	0.35	(0.26)	11.27	(20.42)	23.49	(17.01)	7.47	(19.58)	265
Financially Literate									
yes	0.36	(0.25)	14.13	(20.80)	24.02	(15.98)	8.08	(17.68)	283
no	0.38	(0.26)	11.05	(20.27)	24.00	(17.14)	8.47	(18.09)	277
Stock Owner									
yes	0.45	(0.29)	12.79	(18.20)	22.66	(14.55)	8.95	(13.82)	107
no	0.35	(0.24)	12.50	(21.13)	24.29	(16.99)	8.11	(18.69)	454

"Financially Literate" is an indicator variable which is 1 whenever the respondent states that he/she is either "good" or "very good" with financial matters. For details on this and the other variables, see Appendix 2.A8.

Table 2.2: Experimental Responses in the SOEP by subgroup

and the lab is that the laboratory participants use more bins than the representative respondents.[22] The median number of bins that contain at least one brick is 6 in the laboratory while it is only 3 in the SOEP where 28% of respondents use only a single bin and a further 14% only use two bins.[23]

In the analysis below we repeatedly use summary statistics that we compute from the reported histograms. To compute statistics like the expectation or the standard deviation of returns from the underlying belief distribution we take the 8 points on the CDF, interpolate between them using a cubic spline and then calculate the statistics numerically.[24] Using these imputed distributions, we find that the average of the SOEP respondents' mean expected return of the fund is 12.5% and the average standard deviation of the fund's return distribution is 24.0%. For the laboratory sample, the average mean belief about the fund's return is 11.6% and the average standard deviation is 35.6%.

As described in the previous section, we also elicited scalar belief reports by asking for the "expected" fund return. In the SOEP sample, this variable has a mean of 8.3% and a standard deviation of 17.8%. In the laboratory sample, the mean is 11.0% and the standard deviation is 19.1%. Stated expectations are highly correlated with expectations inferred from belief distributions (Pearson correlation coefficient: 0.5 for the SOEP and 0.31 for the lab sample). Table 2.2 collects key descriptives for the main experimental variables for different subgroups of the SOEP sample (a similar table for the lab sample is omitted because the student population is very demographically homogeneous).

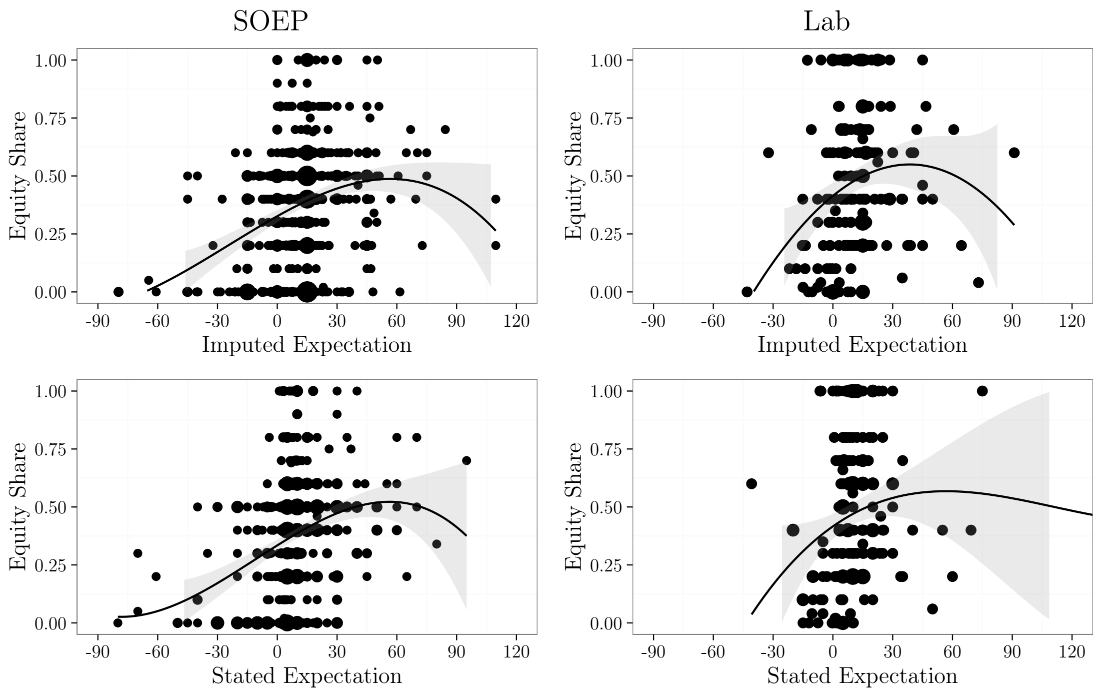
We now investigate the extent to which equity share and beliefs are correlated. Figure 2.1 contains a scatter plot of equity shares and the belief measures for both the SOEP and the lab sample. The figure shows pronounced positive relationships between belief and investment overall. At the mean of the data an increase in the expected return by one percentage point is associated with a one third percentage point increase in the equity share (see Table 2.A3 in the Appendix for OLS re-

[22] Appendix 2.A6 contains examples of the raw data of elicited histograms from both samples.

[23] Relative to comparable studies that use similar methods, the mentioned frequencies are on the low side. Delavande and Rohwedder (2008) report that 73% of their subjects used two or fewer bins.

[24] This method is due to Bellemare, Bissonnette, and Kröger (2012). A more detailed description of the interpolation procedure can be found in Appendix 2.A7.

gressions). This relationship holds for both our belief measures and is roughly the same in the laboratory. This evidence of a positive association between beliefs and investments is consistent with many studies in the belief elicitation literature (see, for example, Naef and Schupp (2009) and Costa-Gomes, Huck, and Weizsäcker (2014) in the context of trust games).



Overlapping observations are aggregated, with the dot's size being proportional to the number of observations thus aggregated. Model fit comes from a polynomial regression in which investments are a cubic function of expected return (Models 2, 5, 8 and 10 in Table 2.A3 in the Appendix, which also contains alternative specification that e.g. control for personal characteristics but all show results that are qualitatively and quantitatively similar.). 95% confidence interval in light gray.

Figure 2.1: Equity Share and Beliefs

Notice that the data also show patterns that are hard to square with the predictions of the standard model. As in Merkle and Weber (2014) there is a substantial fraction of participants who expect a negative excess return for the experimental asset and yet invest positive amounts. But altogether, the statistical connection between belief data and investment decisions can be regarded as supporting the basic implication of the standard portfolio choice model: higher expected returns occur together with larger investments.

Stock-market participation rate by...	Decile									
	1st	2nd	3th	4th	5th	6th	7th	8th	9th	10th
Household Income	7%	7%	3%	21%	14%	17%	20%	19%	26%	46%
Liquid Wealth	0%	2%	2%	2%	5%	13%	11%	39%	43%	56%

Table 2.3: Stock-market participation rate by income and wealth deciles

2.3 External validity: Stock market participation

We now turn to the important question whether our response variables are indicative of real-life investments. Specifically, we test the external validity of our data by comparing elicited behavior in the experiment with survey responses to the question "Do you own any stock market mutual funds, stocks or reverse convertible bonds ("Aktienanleihen")?"

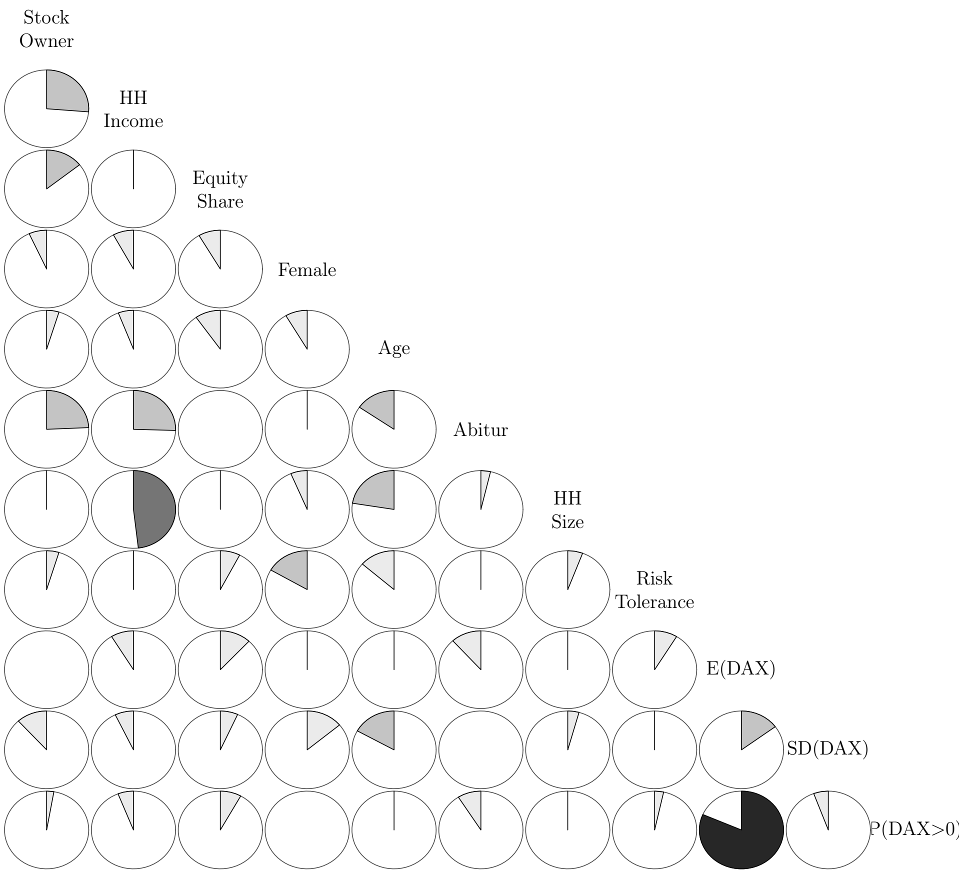
18% of all households answered this question in the affirmative, which is in line with other evidence on the German stock market participation.[25] Splitting the participation rate by deciles of both household income and a proxy for liquid wealth[26] Table 2.3 also shows that stock market participation increases in both variables but stays well below 100%.

Figure 2.2 displays a correlogram, a visualization of the correlation matrix for several survey and experimental variables. Starting from the vertical, positive correlations are displayed as wedges that are shaded clockwise while negative correlations are shaded counter-clockwise. The higher the correlation, the larger the wedge and the darker the shade of the wedge.

The correlogram shows that only a handful of variables are reliable predictors of stock market participation. Most of the significant correlations have been observed in the previous literature. For example, household size is known to be a significant

[25] Most other surveys provide numbers only for the percentage of individuals who hold stocks. In our data this percentage stands at 15.4% (S.E.: 1.1%) while a 2012 survey by Deutsches Aktieninstitut (2012) puts it at 13.7%.

[26] The SOEP question about interest earned on investments over the previous year is answered by far more people than more detailed questions about the amounts of wealth held in the form of various assets. We therefore use this variable as a proxy for liquid wealth. The alternative measure, the sum over all asset classes, yields broadly similar results. For details on these variables, see Appendix 2.A8.



The correlogram above visualizes the pairwise (Pearson) correlation coefficients of the variables. E(DAX) is the imputed expected return on the DAX going forward while SD(DAX) is the imputed standard deviation of the reported return distribution. P(DAX>0) is the reported probability that the DAX will make a gain over the next year.

Figure 2.2: Correlogram

correlate of stock market holdings. Likewise, household income and *Abitur*—the highest form of secondary education in Germany and the only form that grants access to the university system—are well-known and entirely unsurprising predictors of stock ownership. Notice that equity share is the only experimental variable that has predictive power for stock holdings (correlation: 0.14, p-value: < 0.001).

Of course, the correlograms only show bivariate relations. In order to gain a broader picture we investigate whether the correlations change if we control for other variables.[27] We find that equity share has explanatory power over and

[27] This is similar to the approach taken by Guiso et al. (2008) who study the co-variation of stock

above the other variables, see Table 2.4. Even after including all relevant controls, which drives up the R^2 to around 30%, the coefficient for equity share remains both economically important and statistically significant and is robust to different specifications. Back-of-the-envelope calculations yield the result mentioned in the introduction, that an increase in equity share by one standard deviation is associated with an increase in stock market participation of six percentage points.

The fact that equity share helps to explain stock holdings even if we control for all other variables that are known to be good predictors of stock market participation is important for two reasons. First, it establishes external validity. Investment behavior in the experiment is strongly related to investment behavior outside of the experiment. Second, the result gives hope that the simple experimental portfolio choice problem can be used as a wind tunnel: it allows the controlled manipulation of a behavioral variable that has a close connection to stock market particpation, both in terms of economic theory and in terms of empirical correlation. Hence, there is hope that interventions, for example, to encourage stock ownership, could be pre-tested in laboratory or artefactual field experiments such as ours.

market participation with generalized trust and other variables.

	Dependent variable: Stock Market Participant		
	(1)	(2)	(3)
Equity Share	0.220***	0.240***	0.200***
	(0.072)	(0.068)	(0.064)
Female		−0.043	−0.029
		(0.032)	(0.030)
Born in East Germany		−0.058*	−0.044
		(0.034)	(0.033)
Age		0.006	0.004
		(0.005)	(0.006)
Age^2		−0.0001	−0.0001
		(0.0001)	(0.0001)
Abitur		0.200***	0.150**
		(0.061)	(0.058)
University Degree		0.049	−0.003
		(0.078)	(0.072)
Household Size		0.039**	−0.004
		(0.019)	(0.022)
Risk Tolerance: Low		0.020	0.034
		(0.037)	(0.035)
Risk Tolerance: High		0.008	0.058
		(0.044)	(0.043)
Imputed expectation of DAX		0.001	0.0003
		(0.001)	(0.001)
Imputed S.D. of DAX		−0.003***	−0.001
		(0.001)	(0.001)
Gain Probability of DAX		−0.003	0.039
		(0.088)	(0.085)
Number of Children in Household		−0.096***	−0.057*
		(0.030)	(0.030)
Employed		−0.015	−0.024
		(0.036)	(0.037)
Financially Literate		0.140***	0.080***
		(0.032)	(0.031)
Interest: < 250 Euros			0.061*
			(0.033)
Interest: 250 - 1.000 Euros			0.270***
			(0.057)
Interest: 1.000 - 2.500 Euros			0.430***
			(0.086)
Interest: > 2.500 Euros			0.310***
			(0.110)
Interest: refused to answer			0.150
			(0.100)
Household Income (missing=0)			0.023
			(0.018)
Household Income: missing			0.210**
			(0.084)
Constant	0.110***	−0.130	−0.130
	(0.029)	(0.140)	(0.140)
N	561	560	560
R^2	0.021	0.150	0.280
Adjusted R^2	0.019	0.130	0.250

*p < .1; **p < .05; ***p < .01
Household income is in thousands of Euros

Household income is set to zero where missing (48 cases). Moreover, a dummy variable is added to the regression which is 1 for the observations with missing household income.

Table 2.4: Predicting real-world stock-market participation

2.4 Treatment effects

Recall that we implement five exogenous treatments that shift the historical return of the DAX. The shifts are sizable, ranging from -10 percentage points to +10 percentage points. Table 2.5 documents that by and large there is, surprisingly, no effect of the return shifter on equity share in the SOEP sample. The lack of response can hardly be explained by small incentives. In terms of the nominal framing of the €50,000 investment, the difference in returns between Treatments -10 and 10 amounts to a difference in returns of up to €10,000. In terms of the real monetary value of the experimental investment, the variation in return amounts to a difference of up to €5. This difference is large enough for the typical participant in an experiment (even in representative samples) to react. The overall lack of response therefore suggests that many respondents find it difficult to incorporate the shift appropriately in their investment choice.

However, this result is not universal. Instead we notice an important difference between the SOEP and the laboratory sample. While SOEP participants appear to ignore the shifter on average, there is a strong and statistically significant reaction of investments to the treatment in the laboratory. There, the equity share rises from 0.30 to 0.63 in response to improving the return of the fund by 20 percentage points.

Similar results hold for those parts of the SOEP sample that are plausibly more financially savvy, those who are more educated, those who have more liquid assets (or refuse to answer the question about how much interest they obtain from liquid assets) and those who answer the standard financial literacy question about compound interest correctly. Hence, it appears that the main difference between SOEP and lab is driven by selection on educational covariates and wealth.[28]

The beliefs about the fund's return, however, do not respond to the shifter in the way they should, no matter what measure of beliefs we use and no matter whether we consider the SOEP data or the laboratory data and no matter how we slice the data. While there is a statistically significant effect in the laboratory sample, it is much smaller than the 20 percentage points predicted by probabilistic

[28] For details of differences between subsamples, see Appendix 2.A3.

Setting	Variable	-10	-5	0	5	10	ANOVA	Kruskall-Wallis
SOEP	Equity Share	0.40 (0.02)	0.34 (0.02)	0.32 (0.02)	0.39 (0.02)	0.39 (0.02)	0.106	0.135
	Imputed Beliefs	13.14 (1.97)	10.58 (1.81)	9.38 (1.85)	14.48 (1.83)	14.45 (2.18)	0.232	0.326
	Stated Beliefs	8.55 (1.71)	7.68 (1.70)	6.60 (1.98)	9.28 (1.43)	8.93 (1.66)	0.810	0.990
	Probability of a Gain	0.68 (0.03)	0.67 (0.03)	0.67 (0.03)	0.74 (0.02)	0.69 (0.03)	0.323	0.313
Lab	Equity Share	0.30 (0.03)				0.63 (0.03)	0.000	0.000
	Imputed Beliefs	10.05 (1.71)				13.37 (1.57)	0.156	0.016
	Stated Beliefs	9.87 (2.28)				12.30 (1.38)	0.374	0.004
	Probability of a Gain	0.59 (0.02)				0.65 (0.01)	0.029	0.009

Table 2.5: Mean levels by treatment

sophistication, and there is no effect at all in the SOEP sample. In both samples and regardless of whether we consider imputed beliefs or stated beliefs, we can strongly reject the rational prediction that the shifter moves the mean of beliefs one-to-one.

We tentatively conclude from this evidence that it is much harder to manipulate beliefs than to elicit them. As we show in Appendix 2.A1 subjects' beliefs about past DAX returns are surprisingly accurate. Within each of the seven histogram bins, the population-average belief of DAX returns falling in the bin is within just few percentage points of the historical frequency. But just like the investments, the beliefs do not react strongly enough to the experimental manipulation. This also raises the question how well the respondents understand the manipulation, despite our long and intense efforts for clarity in the instructions. The next section investigates the possibility that the weak reaction to the manipuation may be driven by factors beyond the understanding of the experimental instructions.

2.5 Asset Complexity and Reactions to Changes in Incentives

In this section we investigate the role of complexity with an additional laboratory experiment. We introduce manipulations of both the risky asset and the safe asset that are economically equivalent and described in identical terms. Yet, the experiment shows that the reaction to an increase in the risky asset's return is weaker than the reaction to an increase in the safe asset's return. This effect is largely consistent with the available evidence on reactions to tax incentives as, e.g., in Chetty et al. (2009) and Abeler and Jäger (2015).

In the additional experiment, the excess return of the risky asset is varied in two ways: either a shift of Δ in the risky asset's return, or a shift of $-\Delta$ in the safe asset's return. To make the two shifts economically equivalent, we modify the decision maker's exogenous income level, as detailed in the next subsection. However, before proceeding to the details, two remarks are in order: First, we designed this section's experiment after we observed the results from the experiments described in Section 2.1.2—hence the separate presentation. Second, the fact that we could run the complexity experiment only in a laboratory format also means that we cannot investigate the present research question for the subsamples that show the weakest reaction to incentive shifts. We suspect, but have no proof, that these subsamples would exhibit even larger differences in their reactions to different shifts.

2.5.1 Experimental Design

The design follows the same format as the paper's main experiment, implementing the standard portfolio choice problem. In the new experiment (i) each participant makes eight investment decisions, allowing a within-subject analysis, and (ii) each participant receives a task-specific fixed income in addition to the earnings from the portfolio choice.

The participants are endowed with an illiquid asset that generates the fixed income W_I, and with liquid wealth W_L that they can allocate among a safe asset and a risky asset. The risky asset pays a rate of return r whereas the safe asset pays a rate of return r_f.

Now consider an increase in the risky return r by an amount Δ, analogous to the exogenous return manipulation of the paper's main experiment. Under this manipulation, a decision maker who invests α in the risky asset earns a random payoff given by:

$$\pi(\alpha) = \alpha W_L(1 + r + \Delta) + (1 - \alpha)W_L(1 + r_f) + W_I$$

For a framing variation of this manipulation by Δ, we can alternatively induce a simultaneous shift in r_f by amount $-\Delta$ and in W_I by amount ΔW_L, yielding the same payoff from investing a share α in the risky asset:

$$\pi(\alpha) = \alpha W_L(1 + r) + (1 - \alpha)W_L(1 + r_f - \Delta) + (W_I + \Delta W_L)$$

From the fact that $\pi(\alpha)$ is identical between both treatments and for all α, we conclude that the same risks are available between the two manipulations. Consequently, expected utility theory, and any other theory that employs a stable mapping from a constant set of uncertainty states into the risky asset's return rate, predict an identical choice by the decision maker. The same statement is true if both the safe and the risky assets' returns are additionally shifted by a constant amount Δ'. The experiment's null hypothesis is thus that participants react equally between the equivalent manipulations of incentives applying to the safe asset or the risky asset.

To ensure that the results are not driven by an asymmetry between positive shifts and negative shifts, we formulate the entire experiment such that only positive shifts occur. This is achieved by adding an appropriate return shift Δ' to both assets.[29] The parameters for the eight choice problems are displayed in Table 2.7. The collection of equivalent variations is the following: Problems 1 and 3 are economically equivalent, Problems 2 and 4 are economically equivalent, Problems 5 and 7 are economically equivalent, and Problems 6 and 8 are economically equivalent. Problems 1 and 2 differ only in the risky asset's return; Problems 3 and 4 differ in the shifter applied to the riskless asset (and a compensatory change in the illiquid endowment), in the described way. But the difference in incentives is the same between 1 and 2 as between 3 and 4. Thus, expected utility and most of its generalizations predict that the difference in investments is identical. Analogously, the difference between 5 and 6 is predicted to be identical to the difference in investments between 7 and 8. As described above, our main hypothesis in this experiment is that shifts in safe return generate a stronger reaction: investments

[29] We also ran three pilot sessions but do not use the data gathered in these sessions here. In the first pilot session subjects were presented with both "bonuses" and "fees" on the two assets and displayed aversive reactions to any asset to which a fee was applied. Since the effect of gain/loss framing was not the subject of this study we therefore ran two sessions with bonuses only but found that up to 42% of subjects chose investments at the lower boundary of the budget set. Since this much truncation presents problems both in terms of power and in terms of the distributional assumptions one is required to make to deal with it, we therefore changed the magnitude of the bonuses to arrive at the valued reported here, values that yield much fewer truncated responses. Note, however, that the responses in all pilots were also indicative of stronger reactions to changes in the safe asset.

Treatment	Bonus on Risky Asset	Bonus on Riskless Asset	Illiquid Endowment	Liquid Endowment
1	9.00	5.90	16000	50000
2	2.65	5.90	16000	50000
3	5.90	2.80	17550	50000
4	5.90	9.15	14375	50000
5	9.10	6.05	14275	50000
6	3.10	6.05	17275	50000
7	6.05	3.00	15800	50000
8	6.05	9.00	15800	50000

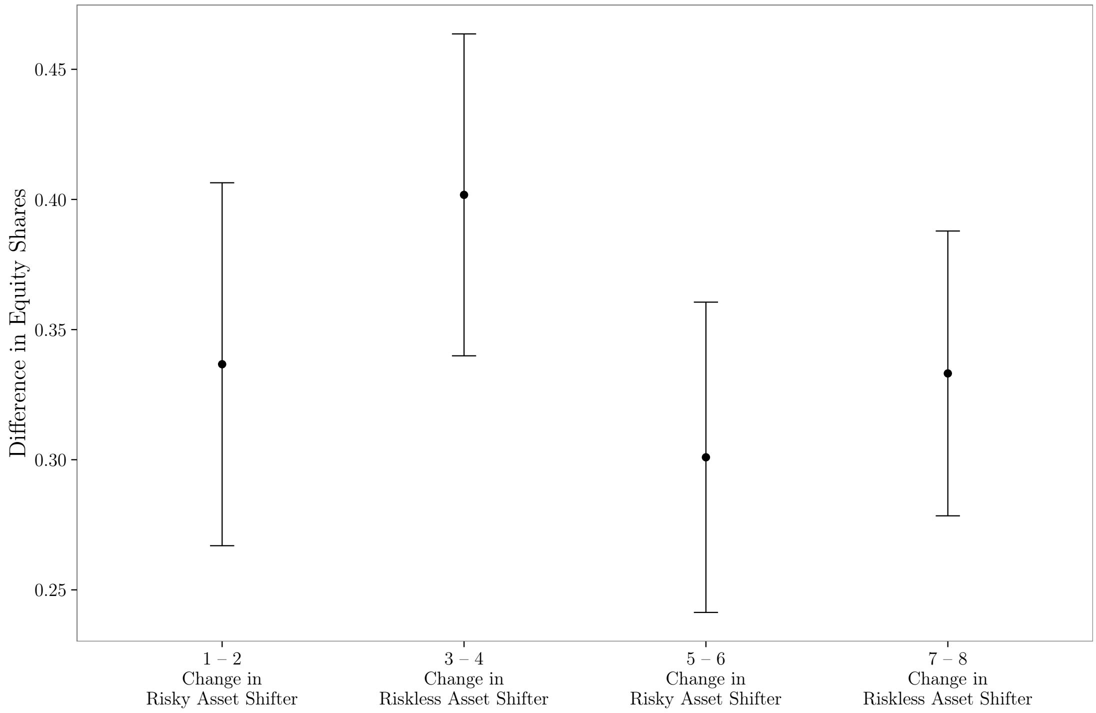
Table 2.7: Treatment parameters

may differ more between 3 and 4 than between 1 and 2, and more between 7 and 8 than between 5 and 6.

76 participants were recruited into 4 experimental sessions at WZB-TU Berlin laboratory in the spring of 2014, using identical procedures as in the study described in Section 2.1.2. Similar to the first lab study we take a fixed-interest German government bond (here, yielding 2 % per annum) as the safe asset and the return on the DAX in a year randomly drawn from 1951 to 2010 as the risky asset. Treatments were presented in random order so as to avoid confounds from learning or contrast effects. One of the eight tasks was randomly selected and paid out at the end of the experiment, ensuring incentive compatibility for each task.

2.5.2 Results

Figure 2.3 displays the differences in average equity shares (the percentage of the liquid endowment invested in the risky asset) for each of the four treatment pairs. A weaker reaction to changes in the risky asset return is immediately visible. Treatments 1 and 2 vary the risky asset return by 6.35 percentage points while holding the riskless asset return constant. This causes a change in mean equity share from 0.28 when the bonus on the risky asset is 2.65 percentage points to 0.62 when the bonus on the risky asset is 9 percentage points for a difference of 0.34. A change of equal magnitude in the return of the riskless asset causes a larger



Error bars show 95% confidence intervals.

Figure 2.3: Investments in the risky asset by treatment

change in the equity share. While the mean equity share in treatment 3 is 0.61, almost identical to that in treatment 1, the mean equity share in treatment 4 is 0.21, lower than that in treatment 2. This yields a difference of 0.4. The same pattern of responses hold analogously for treatments 5 to 8.[30]

Given the comparatively small sample size, each of these mean responses is subject to considerable sampling error. In order to formally test our main hypothesis we therefore pool the data from all treatments. We compute the difference in differences for treatments 1 to 4 and add to this the difference in differences for treatments 5 to 8. Under the null of rational, equal-sized responses to changes in either the risky and riskless asset returns this sum should be zero. Instead, we find it to be 0.10, positive and statistically significantly so (two-sided Wilcoxon rank sum test p-value = 0.03, two-sided t-test p-value = 0.09).[31]

[30] A graph of the raw responses is available in Appendix 2.A10.

[31] Over all treatments about 11% of responses are truncated below at zero. The percentage of

2.6 Conclusion

The paper at hand describes a simple portfolio choice problem with one safe and one risky asset, implemented in an artefactual field experiment for a representative population sample in Germany. The data from this experiment exhibit high degrees of external validity as shown through direct comparison of behavior inside and outside the experiment. This may be viewed as a success for the standard portfolio choice model. Despite its extreme reductionism it captures important real-life tradeoffs in financial markets.

The analysis also shows that the degree of external validity varies between different subgroups. External validity is stronger for skilled and savvy subjects. We also observe that only these savvier subgroups of subjects respond in a meaningful way to changes in incentives, highlighting, once again, the important role of cognitive ability for even the simplest financial decision problems (Benjamin, Brown, & Shapiro, 2013). In our setting less educated subjects forgo substantial additional earnings by not responding to exogenous shifts in investment incentives. Related to previous studies on financial literacy (e.g. Lusardi and Mitchell (2011) on retirement savings, Gerardi, Goette, and Meier (2013) on mortgage foreclosure and von Gaudecker (2015) on portfolio diversification), this difference addresses the possibility of distributional effects that arise from cognitive differences. Similar interventions to foster investments in real life (such as tax subsidies for equity holdings) could have similar undesired effects.

In a separate experiment, we also find evidence that asset complexity is a factor in this under-reaction to incentives. Even university students, who compare favorably with the general population on proxies for cognitive ability, react more strongly to shifts in the return of an asset with a constant return than to shifts in an asset with a stochastic return when both shifts are economically equivalent. To our knowledge, this is a phenomenon that has not yet been documented in the literature on financial literacy, with the exception of the related effects in Chetty et al. (2009) and Abeler and Jäger (2015). This phenomenon—performance in addition

truncated responses is higher in treatments 4 and 8 than it is in treatments 2 and 6. The truncation therefore potentially obscures larger differences between treatments 3 and 4, and 7 and 8, and biases the differences the test statistic towards zero.

can be depend on the nature of the variable to which a number is added—raises questions for the psychology of arithmetic (Ashcraft, 1992) and has potentially numerous applications in the realm of economic decision making under uncertainty.

For future research, our study may inform the design of further wind tunnels for interventions regarding financial investment of households. In particular, in the light of the current underfunding of many pension systems (both pay as you go and capital funded), greater stock market participation by the middle class appears desirable to many economists and policy makers. Testing interventions in artefactual field experiments such as ours might avoid costly mistakes.

2.7 References

Abdellaoui, M., Baillon, A., Placido, L. L., & Wakker, P. P. (2010). The rich domain of uncertainty: source functions and their experimental implementation. *The American Economic Review*, *101* (April), 695–723.

Abeler, J. & Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, *7* (3), 1–28.

Allen, F. (1987). Discovering personal probabilities when utility functions are unknown. *Management Science*, *33* (4), 542–544.

Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2008). Eliciting Risk and Time Preferences. *Econometrica*, *76* (3), 583–618.

Armantier, O. & Treich, N. (2013). Eliciting Beliefs: Proper Scoring Rules, Incentives, Stakes and Hedging. *European Economic Review*, *62*, 17–40.

Arrondel, L., Calvo-Pardo, H., & Tas, D. (2012). *Subjective Return Expectations, Information and Stock Market Participation: Evidence from France.*

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44* (1-2), 75–106.

Bellemare, C., Bissonnette, L., & Kröger, S. (2012). Flexible Approximation of Subjective Expectations Using Probability Questions. *Journal of Business & Economic Statistics*, *30* (1), 125–131.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences. *Journal of the European Economic Association*, *11* (6), 1231–1255.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment. *Quarterly Journal of Economics*, *125* (1), 263–306.

Binswanger, J. & Salm, M. (2013). *Does Everyone Use Probabilities? Intuitive and Rational Decisions about Stockholding* (IZA Discussion Papers No. 7265). Institute for the Study of Labor (IZA).

Bucher-Koenen, T. & Lusardi, A. (2011). *Financial Literacy and Retirement Planning in Germany* (Working Paper No. 17110). National Bureau of Economic Research.

Campbell, J. Y. & Viceira, L. M. (2002). *Strategic Asset Allocation: Portfolio Choice for Long-term Investors*. New York: Oxford University Press.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: theory and evidence. *American Economic Review*, *99*(4), 1145–1177.

Choi, S., Kariv, S., Müller, W., & Silverman, D. (2014). Who Is (More) Rational? *American Economic Review*, *104*(6), 1518–1550.

Costa-Gomes, M., Huck, S., & Weizsäcker, G. (2014). Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect. *Games and Economic Behavior*, *88*, 298–309.

Delavande, A., Giné, X., & McKenzie, D. (2011). Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? *Journal of Applied Econometrics*, *497*(3), 479–497.

Delavande, A. & Rohwedder, S. (2008). Eliciting subjective probabilities in internet surveys. *Public Opinion Quarterly*, *72*(5), 866–891.

Deutsches Aktieninstitut. (2012). *Zahl der Aktienbesitzer geht wieder zurück*. Deutsches Aktieninstitut. Frankfurt am Main.

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2013). Ambiguity Aversion and Household Portfolio Choice: Empirical Evidence. *NBER Working Paper*, (18743).

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes : measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Dominitz, J. & Manski, C. F. (2011). Measuring and interpreting expectations of equity returns. *Journal of Applied Econometrics*, *26*, 352–370.

Dominitz, J. & Manski, C. F. (2007). Expected Equity Returns and Portfolio Choice: Evidence from the Health and Retirement Study. *Journal of the European Economic Association*, *5*(2-3), 369–379.

Forsythe, G. E., Malcolm, M. A., & Moler, C. B. (1977). *Computer Methods for Mathematical Computations*. Englewood Cliffs, NJ: Prentice-Hall.

Fox, C. R. & Tversky, A. (1995). Ambiguity Aversion and Comparative Ignorance. *Quarterly Journal of Economics*, *45*(1), 97–113.

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11267–71.

Goldstein, D. G. & Rothschild, D. (2014). Lay Understanding of Probability Distributions. *Judgment and Decision Making*, *9*(1), 1–14.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125.

Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2011). IQ and Stock Market Participation. *Journal of Finance*, *66*(6), 2121–2164.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *Journal of Finance*, *63*(6), 2557–2600.

Haliassos, M. & Bertaut, C. C. (1995). Why Do So Few Hold Stocks. *Economic Journal*, *105*(September), 1110–1129.

Huck, S. & Müller, W. (2012). Allais for all: Revisiting the Paradox in a Large Representative Sample. *Journal of Risk and Uncertainty*, *44*(3), 261–293.

Huck, S. & Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: results of a lottery choice experiment. *Journal of Economic Psychology*, *20*, 699–715.

Hurd, M. D. & Rohwedder, S. (2012). Stock Price Expectations and Stock Trading. *NBER Working Paper*, (17973).

Hurd, M. D., van Rooij, M., & Winter, J. (2011). Stock market expectations of dutch households. *Journal of Applied Econometrics*, *26*, 416–436.

Hyman, J. (1983). Accurate Monotonicity Preserving Cubic Interpolation. *SIAM Journal on Scientific and Statistical Computing*, *4*(4), 645–654.

Karni, E. & Safra, Z. (1995). The Impossibility of Experimental Elicitation of Subjective Probabilities. *Theory and Decision*, *38*, 313–320.

Kézdi, G. & Willis, R. J. (2009). *Stock market expectations and portfolio choice of american households.*

Lusardi, A. & Mitchell, O. S. (2011). *Financial Literacy and Planning: Implications for Retirement Wellbeing* (Working Paper No. 17078). National Bureau of Economic Research.

Merkle, C. & Weber, M. (2014). Do investors put their money where their mouth is? stock market expectations and investing behavior. *Journal of Banking & Finance*, *46*, 372–386.

Merton, R. (1969). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. *Review of Economics and Statistics*, *51*(3), 247–257.

Naef, M. & Schupp, J. (2009). *Measuring Trust: Experiments and Surveys in Contrast and Combination* (IZA Discussion Papers No. 4087). Institute for the Study of Labor (IZA).

Rabin, M. & Weizsäcker, G. (2009). Narrow Bracketing and Dominated Choices. *American Economic Review*, *99*(4), 1508–1543.

Richter, D. & Schupp, J. (2012). SOEP Innovation Sample (SOEP-IS) — Description and Structure of the SOEP Innovation Sample. *SOEPpapers*, (463).

Rothschild, D. (2012). *Expectations: Point-Estimates, Probability Distributions, Confidence, and Forecasts.*

Samuelson, P. (1969). Lifetime Portfolio Selection by Dynamic Stochastic Programming. *Review of Economics and Statistics*, *51*(3), 239–246.

Stehle, R., Huber, R., & Maier, J. (1996). *Rückberechnung des DAX für die Jahre 1955 bis 1987.*

Stehle, R., Wulff, C., & Richter, Y. (1999). *Die Rendite deutscher Aktien in der Nachkriegszeit - DAX-Rückberechnung für die Jahre 1948 bis 1955.*

Trautmann, S. T. & van de Kuilen, G. (2015). Belief Elicitation: A Horse Race among Truth Serums. *Economic Journal*, *125*(589), 2116–2135.

van Rooij, M., Lusardi, A., & Alessie, R. (2007). *Financial Literacy and Stock Market Participation* (Working Paper No. 13565). National Bureau of Economic Research.

Vissing-Jorgensen, A. (2004). Perspectives on Behavioral Finance: Does "Irrationality"Disappear with Wealth? Evidence from Expectations and Actions. *NBER Macroeconomics Annual 2003*, *18*.

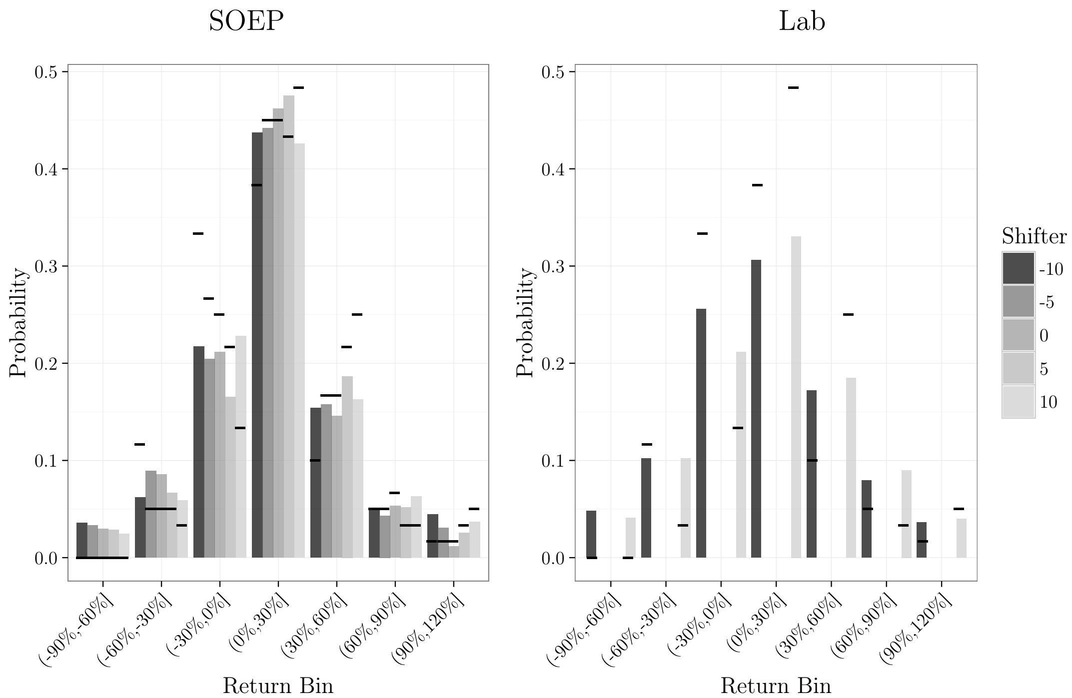
von Gaudecker, H. M., van Soest, A., & Wengström, E. (2011). Heterogeneity in Risky Choice Behavior in a Broad Population. *American Economic Review*, *101*, 664–694.

von Gaudecker, H.-M. (2015). How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice? *Journal of Finance*, *70*(2), 489–507.

Wilcox, N. T. (1993). Lottery choice: incentives, complexity and decision time. *Economic Journal*, *103*(421), 1397–1417.

2.8 Appendices

2.A1 Calibration



Historical benchmark for each treatment indicated by black horizontal lines.

Figure 2.A1: Historical distribution of returns vs. the average distributions in Lab and SOEP

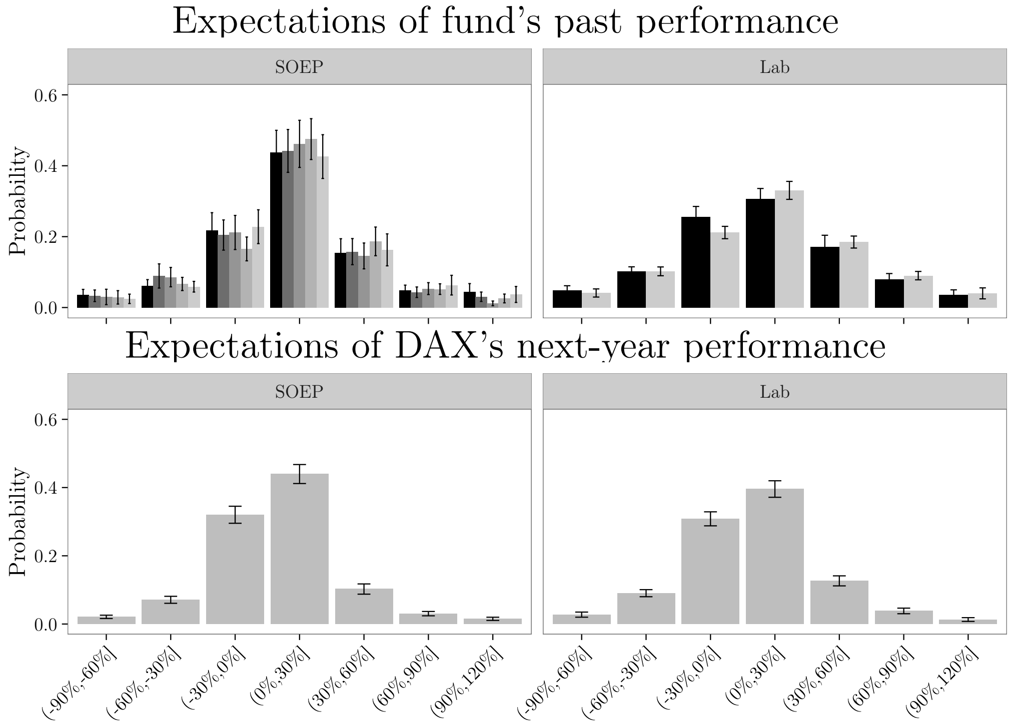
Figure 2.A1 compares the respondents' beliefs about the fund's return with the true historical distribution of DAX returns. The figure shows, in different shades of gray and ordered from left to right within each bin, the five different distributions of beliefs for the five different treatments. The figure also compares these distributions with five corresponding true distributions, indicated by black horizontal lines for each bin and treatment, that result from the true historical distribution plus the five shifters (in the same order, that is, from -10 to the very left to +10 to the very right, within each bin). The figure shows that SOEP respondents are remarkably well calibrated. In none of the seven bins are respondents off by more

than 5 percentage points when data are pooled across treatments. The largest two deviations are that the frequency of small losses between 0 and 30% is slightly underestimated and the frequency of larger losses is slightly overestimated. The good calibration can also be seen in other metrics. While the mean return on the DAX from 1951 to 2010 was 15.5%, both the imputed and the stated expected return on the experimental asset of 12.5% and 8.3% respectively—while lower—are at least similar in magnitude to the historical mean. Moreover, while the relative frequency of a positive return over these six decades was 70.0%, SOEP respondents thought the DAX had seen a gain 69.3% of the time.[32] In contrast, the average distribution of our student subjects in the lab (also shown in Figure 2.A1) differs significantly from the historical benchmark in that too much probability mass is assumed to be in the tails of the distribution.

Underneath the excellent calibration of the average SOEP respondent's belief lies, however, substantial heterogeneity in beliefs and miscalibration at the individual level. Very few of the distributions provided by individual respondents are close to the historical benchmark, and what produces the excellent calibration in the aggregate is a mixture of respondents who put the entire probability mass into a single bin and respondents who report diffuse distributions.

That the return expectations we elicit show such remarkable calibration stands in contrast to evidence from other countries, where substantial miscalibration is commonly observed. For the US Kézdi and Willis (2009) report that HRS respondents expected a stock market gain with roughly 50% probability in the 2002, 2004 and 2006 waves while the historical frequency of a gain on the Dow Jones was 68%. Similarly, the probability of a gain larger than 10% was estimated at 39% but the corresponding frequency was 49%. Dominitz and Manski (2011) find similar numbers in the monthly surveys of the Michigan Survey of Consumers from mid-2002 to mid-2004. In the Netherlands, Hurd et al. (2011) find that in 2004 the median expected rate of return on the Dutch stock market index was a mere

[32] In order to predict whether subjects invest in the risky asset, a relevant question—under expected utility, the *only* relevant question—is whether respondents expect a strictly positive excess return, i.e. a mean return that exceeds 4%. Based on reported beliefs, the proportion of respondents who expect a strictly positive excess return is 69.2% when using stated beliefs, and 72.6% when using imputed beliefs. The historical frequency of the DAX returning strictly more than 4% is 68.3%.



Error bars are 95% confidence interval.

Figure 2.A2: Average distributions of past and future returns

0.3%, a severe underestimate of the historical median return of 14%. A downward bias in expectations is by no means a universal finding, however. Respondents in the 1999, 2000 and 2001 waves of the Survey of Economic Expectations reported expectations for the S&P500 that were substantially above the historical average, but also held the S&P500 to be more volatile than has been the case historically (Dominitz & Manski, 2011).

What explains these differences with the existing literature? One possible explanation is that the papers quoted above compare respondents' expectations about the future with returns realized in the past. A test for correct calibration in this setting then amounts to a joint test of whether subjects hold the historical distribution of returns to be identical to the distribution of returns in the future and, if so, whether they have an accurate picture of the historical distribution. In contrast, we elicit beliefs about the distribution of returns over a well-defined period of time in the past and can test for calibration without auxiliary assumptions.

The beliefs that we elicit about the next 12 months look, however, fairly similar, if somewhat more pessimistic – see Figure 2.A2. This may not be entirely surprising as the survey period was just after the economic crises in parts of Europe had reached their peak intensity. In contrast to expectations about the past, where SOEP respondents and students differed substantially (with the former being more realistic), we find virtually identical expectations about the future between the two samples. The mean imputed return is 12.5% while the probability of a gain on the DAX is thought to be 58.8% on average. 51.8% of subjects state that they expect a return that is higher than 4%.

2.A2 Equity Share and Beliefs – Regressions

	Dependent Variable: Equity Share									
	SOEP: Stated Beliefs			SOEP: Imputed Beliefs			Lab: Stated Beliefs		Lab: Imputed Beliefs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Imputed Expected Return	0.003^{***}	0.005^{***}	0.005^{***}				0.003^{**}	0.007^{***}		
	(0.0005)	(0.001)	(0.001)				(0.001)	(0.003)		
Imputed Expected Return^2		-0.00002^{***}	-0.00001					-0.0001		
		(0.00001)	(0.00001)					(0.0001)		
Imputed Expected Return^3		-0.00000^{**}	-0.00000^{***}					-0.00000		
		(0.00000)	(0.00000)					(0.00000)		
Imputed S.D. of Return			0.001							
			(0.001)							
Probability of a Gain			-0.010							
			(0.037)							
Stated Expected Return				0.004^{***}	0.005^{***}	0.005^{***}			0.002	0.006
				(0.0005)	(0.001)	(0.001)			(0.002)	(0.005)
Stated Expected Return^2					-0.00001	-0.00001				-0.0001
					(0.00001)	(0.00001)				(0.0002)
Stated Expected Return^3					-0.00000	-0.00000				0.00000
					(0.00000)	(0.00000)				(0.00000)
Constant	0.330^{***}	0.330^{***}	0.370^{***}	0.340^{***}	0.330^{***}	0.400^{***}	0.420^{***}	0.410^{***}	0.440^{***}	0.420^{***}
	(0.012)	(0.013)	(0.110)	(0.011)	(0.013)	(0.110)	(0.028)	(0.035)	(0.030)	(0.037)
Personal Controls	No	No	No	No	No	No	No	No	No	No
N	562	562	560	562	562	560	198	198	198	198
R^2	0.074	0.093	0.160	0.081	0.090	0.140	0.031	0.063	0.016	0.038
Adjusted R^2	0.072	0.088	0.120	0.080	0.085	0.100	0.026	0.048	0.011	0.023

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$
Personal controls include dummy variables for gender, being born in the former GDR, having Abitur, having a university education, being employed, having a high self-assessed financial literacy, owning stocks and for each level of our wealth proxy. They also include age and age^2, household size, the number of children in the household and household income
All standard errors are Huber-White heteroskedasticity-robust

Figure 2.A3: Equity Share and Beliefs

2.A3 Different results for different people

In this section we exploit the rich data set on the SOEP respondents in order to study the role of socioeconomic background variables and direct measures or plausible correlates of savviness. As described in Section 2.4, we find strong differences between the SOEP sample and the university student sample regarding the extent to which they react to incentives. This raises the question of whether there is other evidence that "smart", financially savvy respondents react more strongly to variations in incentives. The following analysis confirms the existence of such differences.

We caution that our examination of heterogeneity in the SOEP sample is a "fishing exercise". However, its results are largely in line with what other studies have documented before, namely the fundamental role of cognitive ability for financial decisions making.

Table 2.A2 documents treatment effects on choices and beliefs for different subgroups. It shows that there are small subsamples of the population that do react to incentives. For respondents with a university degree, the coefficients indicate an increase in equity share of one percentage point per one percentage point increase in return. Moving from the worst shifter of -10 to the best shifter of +10, the equity share is predicted to increase by 20 percentage points. This is similar to the effect we observe in the laboratory study with university students where the equity share increases by 33 percentage points. Hence, it appears that the main difference between SOEP and lab is driven by selection on educational covariates.

The results for respondents with different wealth levels are somewhat mixed. For reasons one can only speculate about, the strongest treatment effect is observed for those who withhold information on income from interest. There is also a notable composition effect between the two largest categories: respondents with low but positive levels of income from interest are predicted to increase their equity share by 14 percentage points when we move from the worst to the best shifter. Those without any interest earnings are estimated to exhibit a negative treatment effect.

Among the financial literacy questions we find a heterogeneous treatment effect only for the compound interest question. The other variables that might capture financial literacy do not show significant interactions with the experimental treatment. While the results on financial literacy and wealth are a bit patchy, overall a picture emerges that is familiar from the literature. Even relatively simple investment tasks as the one we have implemented here appear to be cognitively so complex that sensible responses to variations in parameters are shown only by skilled and sophisticated subjects.

An inspection of the two right-hand columns of Table 2.A2 reveals that when it comes to belief manipulation no systematic patterns emerge. Only one of the interactions is statistically significantly different from zero, but only marginally so.

Given that we can identify some subgroups that react better to incentives, it is not far-fetched to presume that we might also be able to detect a stronger external validity of investment levels for these groups. With less noise in behavior inside and presumably outside the laboratory, the measured correlations between the experimental equity share and stock market participation may increase. Table 2.A1 shows the regression-based conditional correlates of stock market participation, separately for different subgroups. Indeed it is the case that "smarter" subsamples show stronger external validity.

	Stock Market Participant			
	All	Abitur	University Degree	Financially Literate
Equity Share	0.200***	0.370**	0.480	0.230**
	(0.064)	(0.180)	(0.300)	(0.110)
Female	−0.029	−0.120	−0.230	−0.049
	(0.030)	(0.110)	(0.150)	(0.052)
Born in East Germany	−0.044	−0.021	−0.160	−0.083
	(0.033)	(0.120)	(0.190)	(0.061)
Age	0.004	−0.028	−0.062	0.002
	(0.006)	(0.023)	(0.044)	(0.011)
Age^2	−0.0001	0.0003	0.001	−0.00004
	(0.0001)	(0.0002)	(0.0005)	(0.0001)
Abitur	0.150**			0.240**
	(0.058)			(0.100)
University Degree	−0.003	−0.002		−0.041
	(0.072)	(0.097)		(0.120)
Household Size	−0.004	0.036	0.045	−0.020
	(0.022)	(0.087)	(0.110)	(0.035)
Risk Tolerance: Low	0.034	−0.015	−0.0003	0.048
	(0.035)	(0.110)	(0.140)	(0.059)
Risk Tolerance: High	0.058	−0.002	0.098	0.058
	(0.043)	(0.160)	(0.240)	(0.064)
Imputed expectation of DAX	0.0003	0.002	0.001	0.001
	(0.001)	(0.007)	(0.010)	(0.003)
S.D. of DAX	−0.001	−0.002	0.002	−0.001
	(0.001)	(0.004)	(0.005)	(0.002)
Gain Probability of DAX	0.039	−0.051	−0.330	0.062
	(0.085)	(0.310)	(0.480)	(0.160)
Number of Children in Household	−0.057*	−0.110	−0.180	−0.062
	(0.030)	(0.110)	(0.150)	(0.049)
Employed	−0.024	0.033	0.022	−0.007
	(0.037)	(0.120)	(0.210)	(0.067)
Financially Literate	0.080***	0.170*	0.200	
	(0.031)	(0.100)	(0.150)	
Interest: < 250 Euros	0.061*	0.047	−0.033	0.086
	(0.033)	(0.110)	(0.170)	(0.054)
Interest: 250 - 1.000 Euros	0.270***	0.330**	0.270	0.320***
	(0.057)	(0.140)	(0.220)	(0.084)
Interest: 1.000 - 2.500 Euros	0.430***	0.560***	0.560**	0.440***
	(0.086)	(0.180)	(0.240)	(0.110)
Interest: > 2.500 Euros	0.310***	0.150	0.013	0.560***
	(0.110)	(0.170)	(0.300)	(0.170)
Interest: refused to answer	0.150	0.350	0.046	0.260
	(0.100)	(0.250)	(0.360)	(0.170)
Household Income (missing=0)	0.023	0.039	0.029	0.010
	(0.018)	(0.040)	(0.059)	(0.029)
Household Income: missing	0.210**	0.150	0.520	0.140
	(0.084)	(0.330)	(0.560)	(0.130)
Constant	−0.130	0.580	1.400	−0.007
	(0.140)	(0.490)	(0.910)	(0.260)
N	560	122	72	283
R^2	0.280	0.360	0.480	0.320
Adjusted R^2	0.250	0.220	0.260	0.260

*p < .1; **p < .05; ***p < .01

Standard errors are Huber-White heteroskedasticity-robust. Household income is set to zero where missing (48 cases). Moreover, a dummy variable is added to the regression which is 1 for the observations with missing household income. “Financially Literate” is an indicator variable which is 1 whenever the respondent states that he/she is either “good” or “very good” with financial matters. For details on this and the other variables, see Appendix 2.A8.

Table 2.A1: Stock market participation by subgroups

	Equity Share				Imputed Expectation of Fund				Stated Expectation of Fund			
	Mean		Treatment Effect		Mean		Treatment Effect		Mean		Treatment Effect	
Education												
< University Degree	0.373	(0.011)	0.000	(0.002)	12.646	(0.922)	0.107	(0.139)	8.649	(0.815)	0.076	(0.113)
University Degree	0.349	(0.033)	0.010^{**}	(0.004)	11.426	(2.619)	0.325	(0.353)	5.586	(2.039)	-0.115	(0.300)
Interest from Wealth												
0	0.368	(0.017)	-0.005^{**}	(0.002)	13.265	(1.572)	0.110	(0.224)	9.012	(1.597)	0.086	(0.214)
< 250 Euros	0.360	(0.019)	0.007^{***}	(0.003)	10.576	(1.344)	0.320	(0.207)	7.759	(1.113)	0.076	(0.163)
250 - 1.000 Euros	0.344	(0.027)	0.001	(0.004)	18.231	(1.758)	-0.123	(0.297)	9.618	(1.569)	-0.247	(0.301)
1.000 - 2.500 Euros	0.422	(0.048)	-0.005	(0.007)	13.582	(3.266)	0.501	(0.518)	7.783	(1.846)	0.011	(0.204)
> 2.500 Euros	0.382	(0.054)	0.004	(0.007)	7.830	(8.722)	-0.653	(1.246)	5.481	(3.307)	0.206	(0.246)
refused to answer	0.339	(0.073)	0.015^{**}	(0.007)	1.971	(8.978)	0.558	(1.030)	3.353	(3.572)	0.543	(0.351)
Financial Literacy: self-assessed												
'good' or 'very good'	0.360	(0.015)	0.002	(0.002)	14.064	(1.231)	0.287	(0.180)	8.047	(1.059)	0.153	(0.153)
'a little' or 'not at all'	0.381	(0.016)	-0.001	(0.002)	11.052	(1.227)	-0.001	(0.183)	8.479	(1.091)	-0.056	(0.147)
Financial Literacy: compound interest												
correct	0.384	(0.014)	0.004^{*}	(0.002)	13.066	(1.157)	0.177	(0.178)	8.741	(0.865)	0.080	(0.117)
incorrect	0.349	(0.018)	-0.003	(0.003)	11.381	(1.415)	0.119	(0.190)	7.701	(1.431)	0.004	(0.213)
don't know	0.365	(0.059)	-0.003	(0.006)	15.608	(3.751)	-0.161	(0.547)	8.560	(4.725)	0.005	(0.533)
Financial Literacy: volatility												
correct	0.400	(0.047)	-0.005	(0.007)	21.056	(4.591)	-0.415	(0.664)	14.726	(4.607)	-0.763	(0.640)
incorrect	0.372	(0.012)	0.001	(0.002)	11.938	(0.906)	0.161	(0.134)	7.911	(0.755)	0.084	(0.102)
don't know	0.301	(0.041)	0.003	(0.006)	11.234	(3.342)	0.556	(0.439)	4.944	(3.744)	0.980^{*}	(0.561)
Stock Owner												
yes	0.448	(0.028)	-0.002	(0.004)	12.828	(1.756)	-0.054	(0.308)	9.280	(1.417)	-0.439^{*}	(0.237)
no	0.353	(0.011)	0.002	(0.002)	12.483	(0.992)	0.185	(0.142)	8.099	(0.878)	0.157	(0.118)

The table shows the results of multivariate regressions in which, for each set of rows, the outcome variables in the columns are regressed on indicator variables for the different levels of the row variables and a variable for the size of the shifter interacted with the different levels of the row variables. "Mean" and "Treatment Effect" therefore correspond to the constants and slope coefficients in bivariate regressions of the column variables on each of the different levels of the row variables. Standard errors for OLS regressions are Huber-White heteroskedasticity-robust.

Table 2.A2: Treatment effect by subgroups

2.A4 Histogram Belief Elicitation Screen

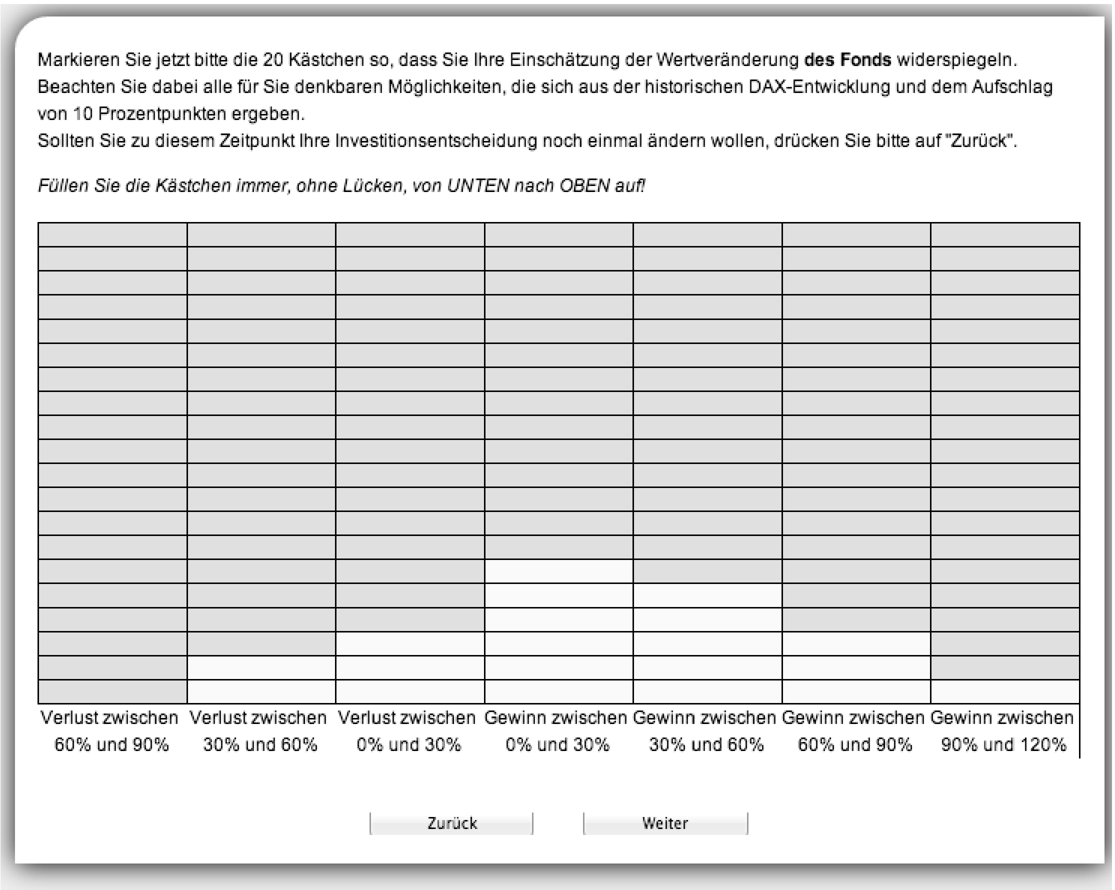


Figure 2.A4: Belief elicitation screen

2.A5 Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Female	700	0.480	0.500	0	1
Age	700	53.000	17.000	16	94
Born in Germany	700	0.860	0.350	0	1
Born in the GDR	700	0.200	0.400	0	1
Abitur	700	0.200	0.400	0	1
University degree	700	0.120	0.330	0	1
Employed	700	0.500	0.500	0	1
Household Size	700	2.300	1.200	1	8
Number of Children in Household	700	0.360	0.780	0	6
Monthly Household Income (in 1000s of Euros)	652	2.500	1.500	0.100	12.000
Risk Tolerance	700	4.900	2.500	0	10
Financial Literacy (self-assessed: 'good' or 'very good')	697	0.500	0.500	0	1
Financial Literacy (compound interest question correct)	690	0.580	0.490	0	1
Financial Literacy (volatility question correct)	690	0.840	0.370	0	1
Equity share (in experiment)	562	0.370	0.260	0.000	1.000
Imputed expectation of fund	562	13.000	21.000	−80.000	110.000
Stated expectation of fund	562	8.300	18.000	−80.000	95.000
Gain Probability of Fund	562	0.690	0.280	0.000	1.000
Imputed expectation of DAX	562	5.500	18.000	−60.000	90.000
Gain Probability of DAX	562	0.590	0.330	0.000	1.000
Total Liquid Assets	515	19.000	44.000	0.000	446.000
Stock Market Participation	693	0.180	0.390	0	1
Stocks (amount)	671	1,780.000	7,874.000	0	110,000
Stocks / Total Liquid Assets	452	0.066	0.190	0.000	1.000
Total Debt	666	17,174.000	54,514.000	0	800,000

N is the number of non-missing observations

Table 2.A3: Descriptive statistics for the 700 heads of household in SOEP sample

2.A6 Some Individual Belief Distributions

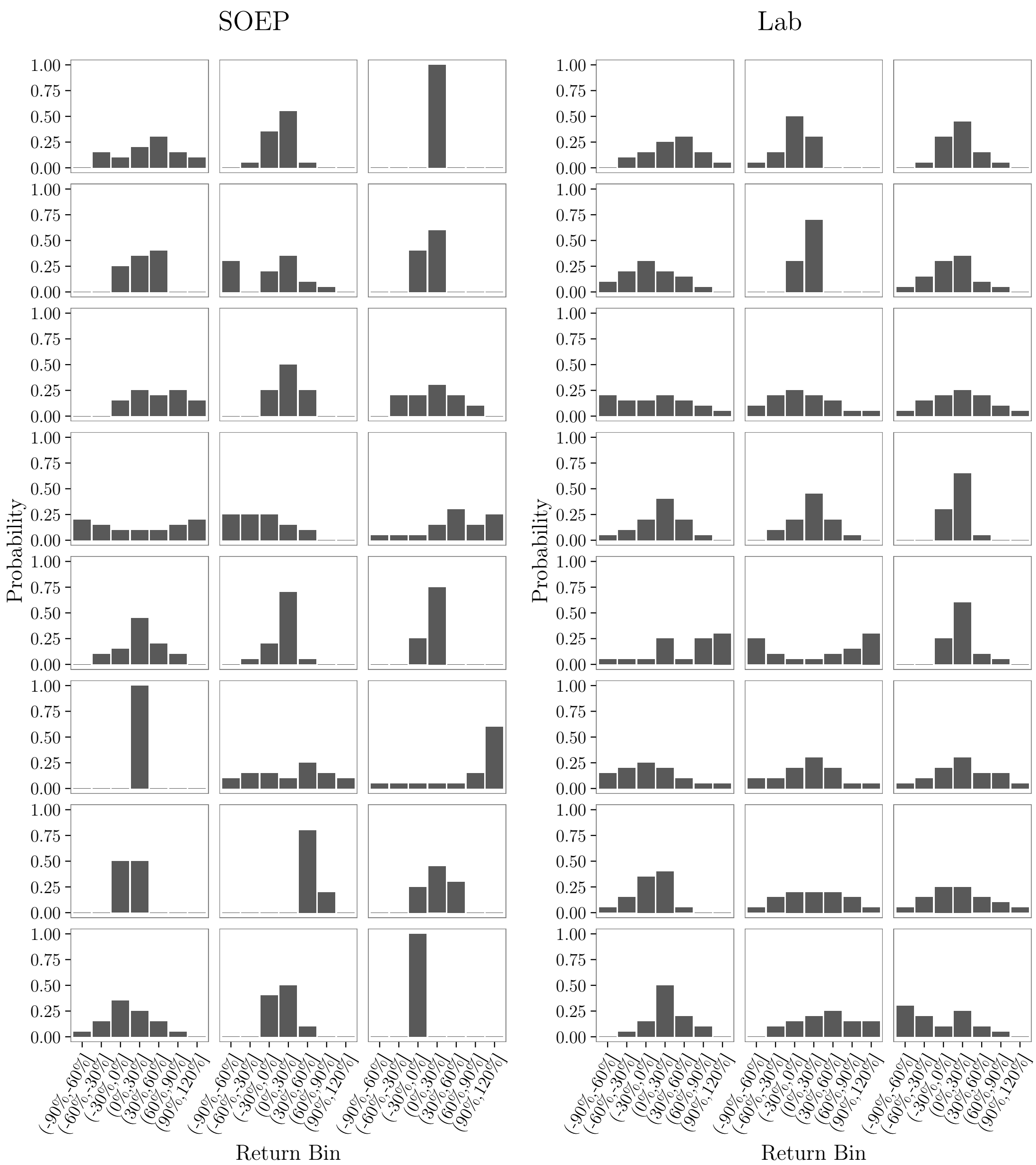


Figure 2.A5: 24 randomly chosen belief distributions from both the SOEP and the lab sample.

2.A7 Imputation of Moments

To derive various summary statistics from the elicited belief distributions we fit continuous distributions to the raw data and calculate the statistics from these distributions.

While much of the existing literature fits parametric distributions we follow an approach similar to Bellemare et al. (2012) and fit cubic interpolating splines using an approach due to Forsythe, Malcolm, and Moler (1977). We first cumulate the probabilities that respondents place within each of the seven bins. This yields 8 points on the cumulative distribution function from which the responses were generated. We take these 8 points to be the knots of the spline (that is, we ignore any rounding in the response and assume that the CDF at these points is *known*) and interpolate between them with a piecewise cubic polynomial.

Since each of the 7 pieces is defined by four polynomial coefficients this is a problem with 28 unknowns. The condition that the spline must go through each of the 8 points gives 14 equations (one each for the end-points and two each for the interior knots) and further assuming that the spline is twice continuously differentiable at each of the knots yields 12 additional equations. What pins down the spline are two boundary conditions, which are found by fitting exact cubics through the four points closest to each boundary and imposing the third derivatives of these cubics at the end-points on the spline.

What is problematic about using such a spline to impute a CDF is that nothing in the procedure described above guarantees that the resulting spline is monotonic. To overcome this problem we apply a filter to the spline that is due to Hyman (1983). The filter relaxes some of the smoothness conditions enough to ensure monotonicity.[33]

Figure 2.A6 demonstrates the fit for six representative respondents. Circles show the raw cumulative probabilities to which both the Hyman-filtered cubic splines as well as various alternative distributions are fitted. By construction the splines are extremely close to the data in all cases – often much closer than any of the parametric distributions that have been fit to the data by minimizing the

[33] Both the Forsythe et al. construction of the spline as well as the Hyman filter are implemented in `R` through the `splinefun()` function with `methods` `fmm` and `hyman` respectively

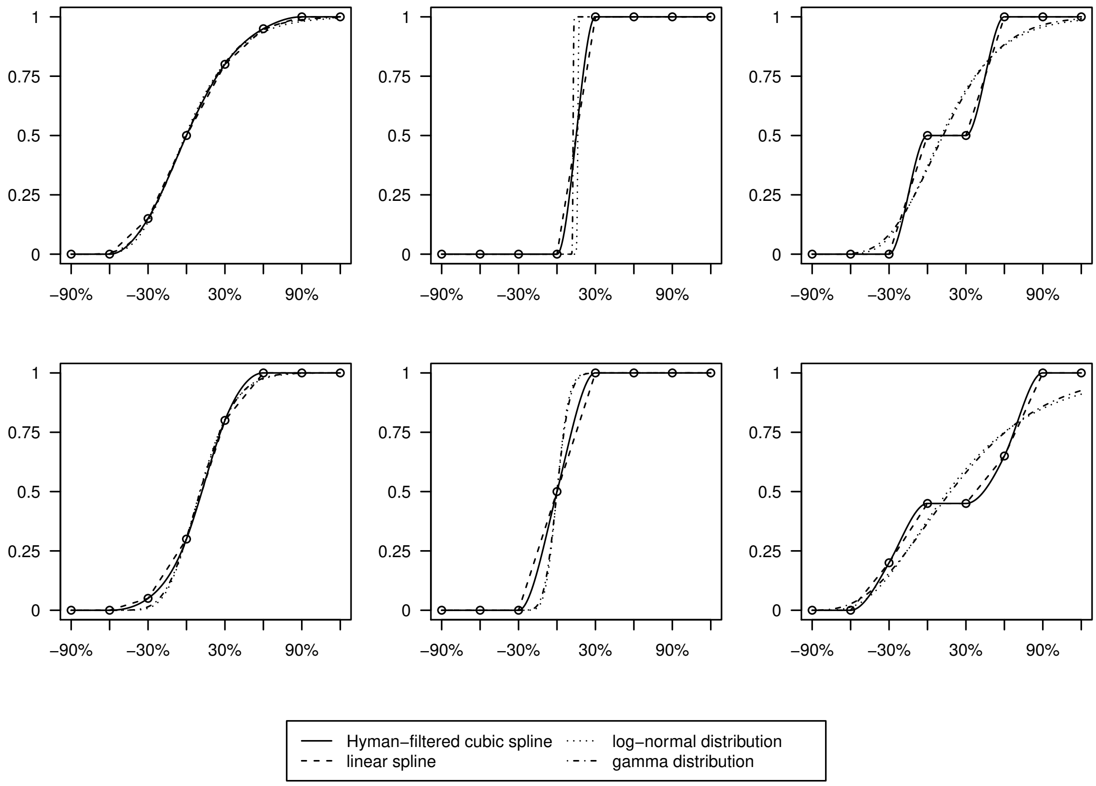


Figure 2.A6: CDFs derived from the belief data using both spline interpolation and parametric distributions fit via least squares

sum of squared deviations at the 8 points. The two distributions on the left are single-peaked and have non-zero probability in several bins and for these cases all of the methods yield roughly the same fit. The distributions in the middle have mass only in a single or in two of the bins, which is a problem for the parametric distributions because in such cases the fit can be improved ad infinitum by reducing the variance of the distribution and thereby reducing the sum of squared deviations at the 8 points. In the two cases on the right the distribution is multi-modal, which naturally leads to terrible fit for the parametric distributions, all of which are unimodal. The splines, in contrast make no such assumptions and therefore fit even these cases rather well.

Finally, we calculate both the mean and the standard deviation from these distributions numerically using adaptive Gauss-Kronrod quadrature.

2.A8 Variable Description and Coding

The full data set contains 1146 respondents in 700 households. Since asset allocation is commonly seen in the literature as the result of joint optimization of all household members we narrow the sample to the 700 heads of household, which we identify as the respondents who filled out the SOEP household questionnaire. All demographics whose coding is detailed below are the demographics of this household head.

Abitur

Germany has a multi-track educational system in which only students who graduate from high school with an "Abitur" diploma are automatically allowed to enroll at university. In the SOEP respondents are asked directly for the highest secondary school degree they have obtained and our Abitur variable is coded mainly according to the answer to this question. There is one special case, however, that requires special attention. 59 respondents obtained their secondary education outside of Germany and a separate question gives too little information to be able to map the secondary education they obtained into the German educational system precisely. Of these subjects, 11 have university degrees, however; education for which, had it been obtained in Germany, the Abitur would almost always be a prerequisite. Since we are interested in the Abitur as a proxy for higher ability and higher education and foreign respondents with university degrees plausibly posses the same higher ability and higher education we recode these subjects as having Abitur.

Born in East Germany

This indicator variable is 1 if the respondent was born in the German Democratic Republic. It is 0 for respondents born in the Federal Republic of Germany, those born outside of Germany and those born in East Germany after reunification in 1990 (14 cases).

Interest from Wealth

This variable is our main proxy for respondents' liquid wealth holdings. Though our survey module included detailed questions about more specific asset classes, item non-response rates for the questions asking for the invested amounts were fairly high. The household questionnaire also included the question "How large, all in all, was your income from interest, dividend payments and capital gains in 2011", with six answer categories.[34] For the econometric analysis we generate a variable that uses information from both questions. We create a new category for subjects who report that their capital income was precisely zero, sort all respondents who gave exact answers into the six categories above and then merged the highest three categories into a single category for capital incomes above €2500 to increase the cell count (counts before the merge were 20 for the €2500 to €5000 category, 5 for the €5000 to €10000 category and 5 for the more than €10000 category). Lastly, we added a category for all subjects who refused to answer both questions.

Financial Literacy

We assess respondents' financial literacy in two different ways. First, we ask people to self-assess their financial literacy with the question:

"**How good, all in all, are you with financial matters?**"[35]

- very good
- good
- a little
- not at all

Second, we ask two questions that explicitly test respondents' financial literacy:

"**Suppose you have €100 in a savings account. You receive 20% on this amount per year and leave the money in the account for 5 years.**

[34] In German: "Wie hoch waren, alles in allem, die Einnahmen aus Zinsen, Dividenden und Gewinnen aus allen Ihren Wertanlagen im Jahr 2011?". Many respondents were either unwilling or unable to provide a precise answer to this question. In a follow-up question they were therefore asked to estimate the amount and choose between 6 categories: below €250, €250 to €1000, €1000 to €2500, €2500 to €5000, €5000 to €10000, more than €10000

[35] In German: "Wie gut kennen Sie sich alles in allem in finanziellen Angelegenheiten aus? Gar nicht, ein bisschen, gut oder sehr gut?"

How much money will be in the account after these 5 years?"[36].

- more than €200
- exactly €200
- less than €200
- don't want to answer

"**Which of the following types of investments has the largest fluctuations in returns over time?"**[37].

- savings accounts
- fixed income securities
- stocks
- don't want to answer

Liquid assets

All household members are asked about individual holdings of the following asset types:

1. checking accounts
2. savings accounts
3. call deposit accounts ("Tagesgeld")
4. fixed deposits
5. covered bonds, municipal bonds, bank bonds, corporate bonds or sovereign bonds
6. stock market mutual funds, stocks or reverse convertible bonds ("Aktienanleihen")
7. real estate funds
8. bond and money market funds
9. other funds
10. other securities

For each of these types, respondents are first asked whether they own any assets of that type at all and, if the question is answered affirmatively, about the size of the

[36] In German: "Angenommen, Sie haben 100 €Guthaben auf Ihrem Sparkonto. Dieses Guthaben wird mit 20% pro Jahr verzinst, und Sie lassen es 5 Jahre auf diesem Konto. Wie viel Guthaben weist Ihr Sparkonto nach 5 Jahren auf?"

[37] In German: "Was glauben Sie: Welche der folgenden Anlageformen zeigt im Laufe der Zeit die höchsten Ertragsschwankungen? Sparbücher, festverzinsliche Wertpapiere oder Aktien?"

asset holdings. Respondents are instructed to estimate this amount should they be unable to provide an exact figure. We code a household as participating in the stock market if the head of household answers the question about stock market mutual funds, individual stocks and reverse convertible bonds with “yes”.

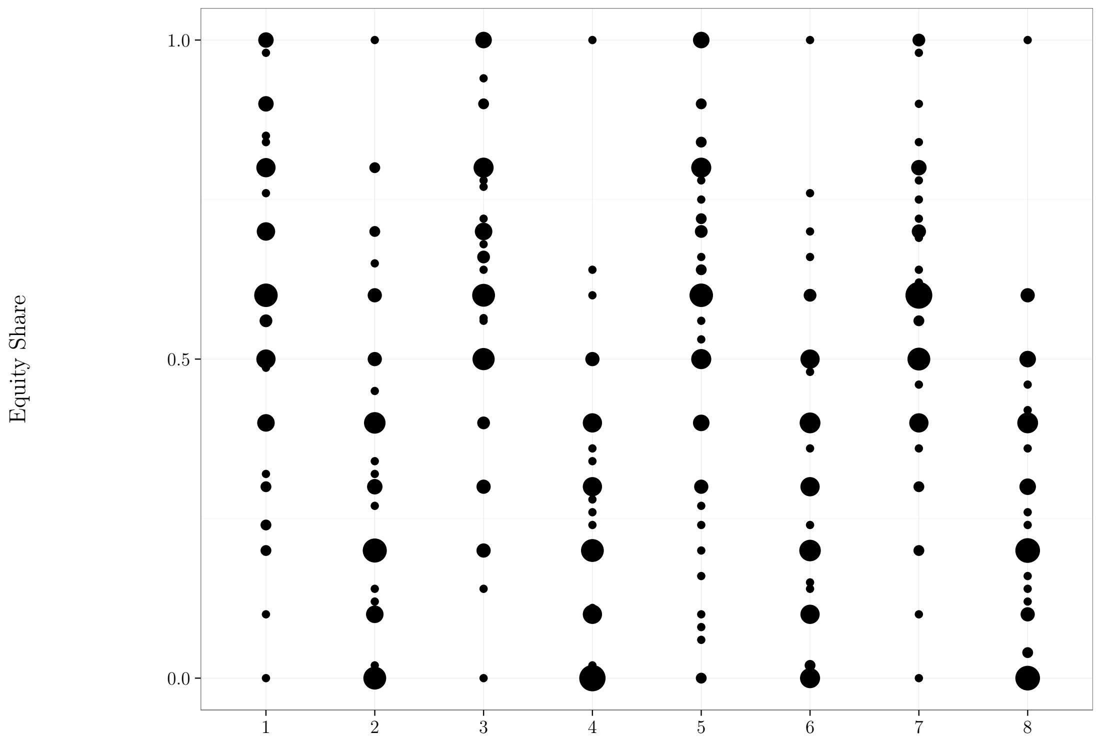
2.A9 Predicting real-world stock-market participation – alternative wealth measures and specifications

	Dependent Variable: Stock Market Participant					
	OLS	OLS	OLS	OLS	OLS	Probit marginal effects
	(1)	(2)	(3)	(4)	(5)	(6)
Equity Share	0.220***	0.240***	0.200***	0.210***	0.140*	0.170***
	(0.072)	(0.068)	(0.064)	(0.066)	(0.076)	(0.056)
Female		−0.043	−0.029	−0.028	−0.028	−0.016
		(0.032)	(0.030)	(0.030)	(0.033)	(0.029)
Born in East Germany		−0.058*	−0.044	−0.032	−0.021	−0.079**
		(0.034)	(0.033)	(0.032)	(0.036)	(0.036)
Age		0.006	0.004	0.002	0.006	0.003
		(0.005)	(0.006)	(0.005)	(0.006)	(0.006)
Age^2		−0.0001	−0.0001	−0.00004	−0.0001	0.000
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.000)
Abitur		0.200***	0.150**	0.140**	0.120*	0.140***
		(0.061)	(0.058)	(0.058)	(0.065)	(0.044)
University Degree		0.049	−0.003	0.013	−0.014	−0.021
		(0.078)	(0.072)	(0.074)	(0.083)	(0.052)
Household Size		0.039**	−0.004	0.003	0.013	0.003
		(0.019)	(0.022)	(0.023)	(0.028)	(0.019)
Risk Tolerance: Low		0.020	0.034	0.033	0.020	0.017
		(0.037)	(0.035)	(0.035)	(0.039)	(0.033)
Risk Tolerance: High		0.008	0.058	0.052	0.058	0.068
		(0.044)	(0.043)	(0.043)	(0.048)	(0.042)
Imputed expectation of DAX		0.001	0.0003	0.0005	−0.0002	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
S.D. of DAX		−0.003***	−0.001	−0.002**	−0.001	−0.002**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gain Probability of DAX		−0.003	0.039	0.035	0.096	0.003
		(0.088)	(0.085)	(0.081)	(0.096)	(0.083)
Number of Children in Household		−0.096***	−0.057*	−0.067**	−0.072**	−0.092***
		(0.030)	(0.030)	(0.031)	(0.035)	(0.031)
Employed		−0.015	−0.024	−0.030	−0.006	−0.015
		(0.036)	(0.037)	(0.037)	(0.042)	(0.039)
Financially Literate		0.140***	0.080***	0.091***	0.078**	0.071**
		(0.032)	(0.031)	(0.032)	(0.036)	(0.030)
Interest: < 250 Euros			0.061*			0.120***
			(0.033)			(0.046)
Interest: 250 - 1.000 Euros			0.270***			0.260***
			(0.057)			(0.047)
Interest: 1.000 - 2.500 Euros			0.430***			0.330***
			(0.086)			(0.058)
Interest: > 2.500 Euros			0.310***			0.270***
			(0.110)			(0.069)
Interest: refused to answer			0.150			0.170*
			(0.100)			(0.090)
Total Liquid Assets (missing=0)				0.011***		
				(0.003)		
Total Liquid Assets^2 (missing=0)				−0.0001**		
				(0.00003)		
Total Liquid Assets^3 (missing=0)				0.00000		
				(0.00000)		
Total Liquid Assets: missing				0.130***		
				(0.040)		
Household Income (missing=0)			0.023	0.032*		0.020*
			(0.018)	(0.017)		(0.012)
Household Income: missing			0.210**	0.230***		0.180***
			(0.084)	(0.082)		(0.069)
Total Liquid Assets					0.012***	
					(0.003)	
Total Liquid Assets^2					−0.0001**	
					(0.00003)	
Total Liquid Assets^3					0.00000	
					(0.00000)	
Household Income					0.020	
					(0.019)	
Constant	0.110***	−0.130	−0.130	−0.100	−0.210	
	(0.029)	(0.140)	(0.140)	(0.130)	(0.140)	
N	561	560	560	560	417	560
R^2	0.021	0.150	0.280	0.290	0.310	

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Income and Liquid assets are in thousands of Euros. Standard errors for OLS regressions are Huber-White heteroskedasticity-robust. Standard errors for probit marginal effects are bootstrapped with 1000 replicates

2.A10 Raw Data in Complexity Experiment



	1	2	3	4	5	6	7	8
Illiquid Endowment	16000.00	16000.00	17550.00	14375.00	14275.00	17275.00	15800.00	15800.00
Bonus on Riskless Asset	5.90	5.90	2.80	9.15	6.05	6.05	3.00	9.00
Bonus on Risky Asset	9.00	2.65	5.90	5.90	9.10	3.10	6.05	6.05

Treatment

Point size is proportional to the number of overlapping observations.

Figure 2.A7: Raw Data in Complexity Experiment

3 Measuring Ambiguity Aversion: Experimental Tests of Subjective Expected Utility

based on work with James Andreoni & Charles Sprenger

3.1 Introduction

Many of the most important economic decisions are made in environments in which no appeal can be made to objective probabilities. In areas such as portfolio allocation, healthcare and insurance coverage, the likelihood of decision-relevant events is not externally "given" but must be judged by the decision maker herself. Subjective Expected Utility (SEU) provides an elegant representation of choice in such contexts (Anscombe & Aumann, 1963; Savage, 1954). Under SEU, a decision maker acts as if she maximizes the expectation of the utility outcomes of her choice with the expectation governed by a coherent subjective probability measure.[1] The Savage development and critical contributions of the Anscombe and Aumann (1963) framework have made SEU a benchmark model of choice throughout economics.

Subjective Expected Utility came under scrutiny from its origination.[2] Notably, the Ellsberg (1961) urn paradoxes drew attention sharply towards potential SEU violations.

Ellsberg's two color problem presents two urns,

[1] That is, the subjective probability of each event is in the interval $[0, 1]$, the probability of the union of disjoint event is the sum of their probabilities, and the entire state space has probability 1

[2] Interestingly, for decision-making with objective probabilities, the famous Allais (1953) paradoxes were distributed at the same conference as Savage's presentation of the Sure-Thing Principle (Savage, 1953) and Samuelson's presentation of Independence (Samuelson, 1953) in May of 1952.

- Urn A: 100 red and black balls in unknown proportion
- Urn K: 50 red balls and 50 black balls,

and requests consideration of the following four choices:

1. Bet red A or Bet black A
2. Bet red K or Bet black K
3. Bet red K or Bet red A
4. Bet black K or Bet black A.

A bet on Urn A is a bet on a subjective event, while a bet on Urn K is a bet on an objective event. A likely choice pattern is indifference in (1) and (2), and strict preference for bets on Urn K in (3) and (4). This violates SEU as there exists no probability measure which can rationalize choice.[3] Originally stated as a thought experiment the Ellsberg two color problem, and variants thereof, have since been demonstrated in laboratory experiments (see Camerer & Weber, 1992, for a survey of the early experimental evidence). Subjective Expected Utility requires a consistency of behavior across subjective event bets that is clearly violated in the data.

Important theoretical contributions have sought to relax the requirement of consistent subjective probabilities. One clarifying development is presented by Gilboa and Schmeidler (1989) who consider decision makers who, when faced with ambiguity, behave as if they did not attribute a single probability measure to an event, but rather consider a set of probabilities and maximize expected utility with respect to the least favorable probability measure from the set. Such a "max-min" decisionmaker can easily display the inconsistency shown in the two color problem. At the most extreme, the decisionmaker is permitted to entertain the pessimistic notions that all the balls in Urn A are black when choosing between the bets in (3) and that all balls in Urn A are red when choosing between the bets in (4). Additional models have been developed with more nuanced objectives of

[3] Under SEU a decision maker strictly prefers a bet on red from Urn K to a bet on red from Urn A iff he believes the probability of drawing a red ball from Urn K to be higher than the probability of drawing a red ball from Urn A. He prefers a bet on black from Urn K to a bet on black from Urn A iff he believes the probability of drawing a black ball from Urn K to be higher than the probability of drawing a black ball from Urn A. Since drawing a red ball and drawing a black ball are complementary events, by additivity and unitarity their probabilities must add to one and so the latter implies that the probability of drawing a red ball from Urn K must be lower than the probability of drawing a red ball from Urn A, which contradicts the belief that rationalizes the first choice.

describing decisionmaking under ambiguity, chief among them the smooth model of Klibanoff, Marinacci, and Mukerji (2005).

Our objective in the present study is to return to the consistency requirements of the SEU model in order to develop deeper insights into individual decisionmaking and potentially provide identification of competing non-SEU models. In addition to consistency à l'Ellsberg between two subjective events, *Subjective-Subjective Consistency*, we consider *Subjective-Objective Consistency*. Subjective-Objective Consistency states that when considering mixtures between subjective and objective bets, a decisionmaker must also behave as if a single probability measure governed her choice.[4] The distinction between Subjective-Subjective Consistency and Subjective-Objective Consistency is important for two reasons. First, of the two popularized non-SEU models we discuss, both deliver Subjective-Subjective inconsistencies while only that of Gilboa and Schmeidler (1989) maintains Subjective-Objective Consistency. Hence investigating Subjective-Objective Consistency allows for separation that a standard Ellsberg experiment cannot deliver. Second, though Ellsberg style experiments are frequent in the literature, questions testing Subjective-Objective Consistency have yet to be implemented experimentally.[5] Particular forms of violations may lead to interesting insights on the structure of decision making when subjective ambiguity can be replaced with objective uncertainty.

We present a design predicated on testing Subjective-Subjective Consistency and Subjective-Objective Consistency. Our design elicits precise valuations of subjective, objective, and mixed subjective-objective bets through a series of binary statements. For example, subjects state whether they prefer a draw from a jar con-

[4] This consistency requirement is closely associated with what is known in the literature as the Certainty Independence axiom. While Anscombe-Aumann Independence requires that preferences not depend on mixing the acts involved with *any* act, Certainty Independence requires merely that preferences be invariant to mixing with constant (objective) acts. Certainty Independence underlies a broad class of non-SEU models, most notably Gilboa and Schmeidler (1989) and Schmeidler (1989). For a full characterization of the class that shares this axiom, the class of invariant biseparable preferences, see Ghirardato, Maccheroni, and Marinacci (2004).

[5] An early literature moved suggestively in this direction by examining situations of partial ambiguity where individuals were told some portion of the distribution from which outcomes would be drawn. These experiments have been infrequent, with some of the treatments in Becker and Brownson (1964), Curley and Yates (1985) and Hong, Bin, and Songfa (2013) being leading examples.

taining an ambiguous composition of 20 red and green marbles, wherein red pays \$10 and green pays \$30, or a draw from a jar of known composition containing 20 yellow and black marbles, wherein yellow pays \$0 and blacks pays \$30. Variation in the jar of known composition that induces a change in preference from the ambiguous jar to the known jar provides a measure of valuation. Note this is effectively the experimental implementation of the Schmeidler (1989) 'risk equivalent' and is the natural extension to the subjective domain of eliciting a cardinal utility index for an objective lottery.[6] Comparing valuations for two complementary subjective bets allows identification of Subjective-Subjective Consistency à l'Ellsberg.[7] Mixtures of subjective and objective bets allow identification of Subjective-Objective Consistency.[8]

In an experiment conducted with 133 subjects at the University of California, San Diego, we document substantial deviations for both of our consistency notions. 79% of subjects violate Subjective-Subjective Consistency largely in a manner reminiscent of Ellsberg's two-color problem. The mean data place valuations of an ambiguous jar 10% below that of a known 50-50 jar regardless of payment labels. Such behavior is rationalized in our data by estimates of beliefs indicating subjects believe an ambiguous jar is composed of around 70% low-paying marbles regardless of the color that pays the low outcome. Surprisingly, over 95% of subjects violate Subjective-Objective Consistency. The mean behavior indicates a "directed pessimism," consistent with subjects believing that an ambiguous jar consists of around 90% low-paying marbles when mixing with a known high outcome, while believing that the same ambiguous jar consists of only around 55% low-paying marbles when mixing with a known low outcome. Interestingly, the correlation between inconsistencies at the individual level is limited, suggesting possible differentiation in individuals' ambiguity attitudes and their treatment of situations in which subjective ambiguity is replaced with objective uncertainty.

[6] A standard cardinal utility index constructed in proofs of von Neumann-Morgenstern expected utility identifies the cardinal utility of a gamble as the mixture of the best and worst options in the space of outcomes that yields indifference. See, for example, Varian (1992).

[7] That is, we compare the valuations of ambiguous urns with either red pays \$10, green pays \$30 or red pays \$30, green pays \$10 relative to a jar with known 50-50 composition.

[8] For example, we examine the valuation of 10 ambiguous marbles and 10 red marbles with red pays \$10, green pays \$30 to the valuation of 10 ambiguous marbles and 10 green marbles with red pays \$10, green pays \$30. The operationalization is described in Section 3.3.

Our exploration of consistency yields several conclusions and insights for future work. Our pattern of deviations is at odds with SEU as well as popularized non-SEU models. Our data, however, give instruction as to what modeling constructs may prove useful in capturing behavior. Potential suggestions include mixture dependence for ambiguity attitude and explicit subjective bet considerations costs. Naturally, however, a full elaboration and foundation for such constructs is outside the scope of this work. Even without an articulated model, our findings may prove useful for the investigation of other behaviors. We consider several potential applications related to the newer finding of Subjective-Objective inconsistency.

The paper proceeds as follows. Section 3.2 presents a conceptual framework for our SEU consistency tests and briefly considers the predictions of alternative models. Section 3.3 describes the design and procedures. Section 3.4 presents results and Section 3.5 provides discussion and a conclusion.

3.2 Conceptual Framework

Our broad setting is that of Anscombe and Aumann (1963). There is a finite state space, S, a set of consequences, X, a set of objective probability distributions over these consequences, $\Delta(X)$, and a set of acts, F, functions which map S into $\Delta(X)$.[9] Individuals carry preferences, $\succsim$, over acts in F.

If preferences over acts satisfy axioms of completeness, transitivity, monotonicity, independence, continuity and non-degeneracy, they can be represented by the Subjective Expected Utility (SEU) functional. For a given act, a,

$$U_{SEU}(a) = \int_S p(s)\ u(a(s))\ ds$$

where $p(s)$ refers to the subjective probability associated with state $s \in S$, $\int_S p(s)ds = 1$ and $u(a(s))$ is the (von Neumann-Morgenstern) expected utility of the lottery that act a yields in state s.

[9] This, rather than the original formulation by Savage (1954) in which states map directly into consequences, has become the canonical choice domain on which models of choice under uncertainty are formulated. Note that objective lotteries are primitives of the model allowing one to describe “roulette” lotteries with probabilities objectively given and “horse race” lotteries for which probabilities are wholly subjective.

With this foundation, we introduce an Ellsberg style jar composed of red and green marbles with the color of a drawn ball determining the state of the world. Hence, the state space is binary, $S = \{red, green\}$. We establish the set of consequences as $X = \{0, x, y\}$ with $0 < x < y$. We consider two simple subjective acts that are complements:[10]

- f: yields y with probability 1 if $s = red$, yields x with probability 1 if $s = green$.
- g: yields y with probability 1 if $s = green$, yields x with probability 1 if $s = red$.

The act f represents a bet on the subjective event *red*, while the act g represents a bet on the subjective event *green*.

In order to elaborate our consistency tests it will be helpful to introduce two degenerate objective acts:[11]

- h: yields y with probability 1 if $s = red$, yields y with probability 1 if $s = green$.
- l: yields x with probability 1 if $s = red$, yields x with probability 1 if $s = green$.

These acts yield the high prize, y, with probability 1 in both states, or the low prize, x, with probability 1 in both states.

Our consistency implications are constructed from these four acts. The development is aided by recalling a key axiom used in the SEU construction, Independence.

Definition 1 (Independence)**.** *For any three acts $a, b, c \in F$ and any $\alpha \in (0, 1)$*

$$a \succsim b \iff \alpha a + (1 - \alpha)c \succsim \alpha b + (1 - \alpha)c. \tag{3.1}$$

Independence states that the preference between two acts should not change when both acts are mixed with any other in equal proportion.[12] Independence helps ensure the SEU formulation is linear in subjective probabilities and that

[10] By simple act we mean that the distribution over X, $\Delta(X)$, induced in each state is degenerate. In the literature these are also known as "Savage acts."

[11] That is, the acts are governed by objective probabilities and induce the same degenerate distribution over X in every state.

[12] Where mixtures are probabilistic mixtures which are performed state-wise. The mixture $\alpha a + (1 - \alpha)c$ denotes an act, which yields the lotteries $\alpha a(s) + (1 - \alpha)c(s)$ in every state s.

the probabilities indeed constitute a proper probability measure. Our consistency implications build closely from the requirements of a coherent probability measure and linearity in probabilities.

3.2.1 Subjective-Subjective Consistency

SEU implies the existence of coherent subjective probability measure between complementary subjective acts. That is, for acts f and g a single subjective probability, $\tilde{p}_{green} = 1 - \tilde{p}_{red}$, must rationalize the valuations for both f and g. Consider the acts f, g and $\frac{1}{2}h + \frac{1}{2}l$, the acts corresponding to Ellsberg's classic 2-color problem. Under SEU these acts are associated with the expected utilities

$$\begin{aligned}
U_{SEU}(f) &= (1 - \tilde{p}_{green}) \cdot u(y) + \tilde{p}_{green} \cdot u(x) \\
U_{SEU}(g) &= \tilde{p}_{green} \cdot u(y) + (1 - \tilde{p}_{green}) \cdot u(x) \\
U_{SEU}(\frac{1}{2}h + \frac{1}{2}l) &= \frac{1}{2} \cdot u(y) + \frac{1}{2} \cdot u(x)
\end{aligned}$$

We present *Subjective-Subjective Consistency* as the restriction on the valuations of f and g relative to the valuation of $\frac{1}{2}h + \frac{1}{2}l$. The term derives from the restricted pattern of behavior between the subjective probabilities associated with f and g.

Subjective-Subjective Consistency: $\frac{1}{2}[U_{SEU}(f) + U_{SEU}(g)] = U_{SEU}(\frac{1}{2}h + \frac{1}{2}l)$

SEU requires that the average valuation of the subjective acts f and g be equal to the valuation of the objective act $\frac{1}{2}h + \frac{1}{2}l$. The implication is clear from noting that $\frac{1}{2}[U_{SEU}(f) + U_{SEU}(g)] = U_{SEU}(\frac{1}{2}h + \frac{1}{2}l) = \frac{1}{2} \cdot u(y) + \frac{1}{2} \cdot u(x)$. Most Ellsberg style experiments examine and reject an implication of this consistency requirement, namely that the valuations of f and g (weakly) lie on opposite sides of the valuation of $\frac{1}{2}h + \frac{1}{2}l$. The majority of such experiments rely on binary choices or on valuing acts via their certainty equivalents (one recent example using valuations and arriving at the standard conclusion is Halevy (2007)), not on the actual utility valuations, but the often obtained pattern of $f \sim g \prec \frac{1}{2}h + \frac{1}{2}l$ makes clear the inconsistency.

From this consistency test it will be helpful to develop a measure of potential inconsistencies. We define the premium placed on subjective acts over their objective counterpart as a *Subjective Premium* defined as

$$\delta_S = \frac{1}{2}[U(f) + U(g)] - U(\frac{1}{2}h + \frac{1}{2}l).$$

The SEU model predicts $\delta_S = 0$. Note $\delta_S < 0$ for individuals exhibiting the standard pattern of ambiguity aversion and $\delta_S > 0$ for individuals exhibiting the opposite pattern of ambiguity seeking. In our experiment, the subjective premium can be measured precisely, giving a clear sense of the size of potential violations.

3.2.2 Subjective-Objective Consistency

A second notion of consistency arises when considering the evaluation of a single subjective act, say g, and altering the decision maker's exposure to this act by mixing with objective acts like h and l. This notion of consistency relies tightly on linearity in mixture proportions.

Consider the subjective act g, with

$$U_{SEU}(g) = \tilde{p}_{green} \cdot u(y) + (1 - \tilde{p}_{green}) \cdot u(x).$$

Combine g with h in proportion $(\alpha, 1-\alpha)$ to arrive at

$$U_{SEU}(\alpha g + (1-\alpha)h) = \alpha \cdot (\tilde{p}_{green} \cdot u(y) + (1-\tilde{p}_{green}) \cdot u(x)) + (1-\alpha) \cdot (1 \cdot u(y) + 0 \cdot u(x))$$

Combine g with l in proportion $(\alpha, 1-\alpha)$ to arrive at

$$U_{SEU}(\alpha g + (1-\alpha)l) = \alpha \cdot (\tilde{p}_{green} \cdot u(y) + (1-\tilde{p}_{green}) \cdot u(x)) + (1-\alpha) \cdot (0 \cdot u(y) + 1 \cdot u(x))$$

This yields two facts about mixtures of a single subjective act with objective acts. First, the valuation of a mixture is linear in the mixture proportion α. Second, the distance between $U_{SEU}(\alpha g + (1-\alpha)l)$ and $U_{SEU}(\alpha g + (1-\alpha)h)$ is independent of the subjective probability, $\tilde{p}_{green}$. That is, when mixing with high constants or low constants an individual values the remaining subjective portion the same way. Hence, differences in valuations between mixtures must be independent of subjective probabilities. To see this consider

$$
\begin{aligned}
U_{SEU}(\alpha g + (1-\alpha)h) - U_{SEU}(\alpha g + (1-\alpha)l) &= (1-\alpha)u(y) - (1-\alpha)u(x) \\
&= u(y) - (\alpha u(y) + (1-\alpha)u(x)),
\end{aligned}
$$

which is identical to

$$U_{SEU}(\alpha g + (1-\alpha)h) - U_{SEU}(\alpha g + (1-\alpha)l) = U_{SEU}(h) - U_{SEU}(\alpha h + (1-\alpha)l).$$

We state *Subjective-Objective Consistency* as the restrictions placed on the above mixture valuations under SEU. The term derives from the fact that only a given subjective act is considered and consistency is required for mixtures with objective acts.

Subjective-Objective Consistency: For a given subjective act and α-proportion mixtures with h and l, the difference in valuations is independent of subjective probabilities.

$U_{SEU}(\alpha f + (1-\alpha)h) - U_{SEU}(\alpha f + (1-\alpha)l) = U_{SEU}(h) - U_{SEU}(\alpha h + (1-\alpha)l)$,

and

$U_{SEU}(\alpha g + (1-\alpha)h) - U_{SEU}(\alpha g + (1-\alpha)l) = U_{SEU}(h) - U_{SEU}(\alpha h + (1-\alpha)l)$.

From this second consistency implication it will be helpful to develop another measure of potential inconsistencies. We define the *Mixture Distances*, $\delta_M(\alpha, f)$ and $\delta_M(\alpha, g)$, as the difference in the distance between the valuations of subjective-objective mixtures and the distance between corresponding purely objective mixture valuations. For example,

$$\delta_M(\alpha, g) = [U(\alpha g + (1-\alpha)h) - U(\alpha g + (1-\alpha)l)] - [U(h) - U(\alpha h + (1-\alpha)l)].$$ [13]

Note the SEU model predicts $\delta_M(\alpha, g) = 0$ and $\delta_M(\alpha, f) = 0$ for all α. Importantly, a non-SEU decision maker who behaves as if a single probability governs

[13] Correspondingly, $\delta_M(\alpha, f) = [U(\alpha f + (1-\alpha)h) - U(\alpha f + (1-\alpha)l)] - [U(h) - U(\alpha h + (1-\alpha)l)]$.

her behavior when facing a single subjective act will also exhibit $\delta_M(\alpha, f) = 0$ and $\delta_M(\alpha, g) = 0$ for all α. Deviations from this benchmark can be delivered from non-linearity in mixture proportions with the nature of the non-linearity determining the direction of inconsistency. In our experiment, the mixture distance can be measured precisely giving a clear sense of the size of potential violations.

The requirements of Subjective-Subjective Consistency and Subjective-Objective Consistency both derive largely from SEU's Independence assumption. The two concepts, however, are differentiated by the observation that Subjective-Objective Consistency may hold even when Subjective-Subjective Consistency fails. In the following subsection we highlight the predictions of non-SEU models documenting where and how each model delivers inconsistencies.

3.2.3 Alternative Theories

We focus our attention on two principal generalizations of SEU: Max-min Expected Utility (MEU, Gilboa & Schmeidler, 1989) and the smooth ambiguity model of Klibanoff et al. (2005, KMM). We present these formulations and describe the potential violations of the SEU consistency requirements that each can accommodate.

Max-min Expected Utility

MEU relaxes the Anscombe-Aumann Independence axiom for subjective acts, using instead

Definition 2 (Certainty Independence). *For all $f, g \in F$, all lotteries r, and for all α in $[0, 1]$*

$$f \succsim g \Longrightarrow \alpha f + (1-\alpha) r \succsim \alpha g + (1-\alpha) r. \tag{3.2}$$

The MEU framework yields a representation in which the probability measure is no longer unique, and decision makers behave as if they maximized expected utility following the most pessimistic (minimal) of these measures. The MEU valuation of an act, a, is

$$U_{MEU}(a) = \min_{p(s) \in P} \int_S p(s)\ u(a(s))\ ds$$

where P is a convex set of probability measures, $\min_{p(s)\in P}$ selects the most pessimistic measure in the set, and everything else is as in SEU. Under our formulation, the act f has valuation

$$U_{MEU}(f) = \min_{\tilde{p}_{green}\in P}(1-\tilde{p}_{green})\cdot u(y) + \tilde{p}_{green}\cdot u(x).$$

The act g has valuation

$$U_{MEU}(g) = \min_{\tilde{p}_{green}\in P}\tilde{p}_{green}\cdot u(y) + (1-\tilde{p}_{green})\cdot u(x).$$

This development helps to clarify which consistency implications will be violated by MEU. First, for Subjective-Subjective Consistency, it is easy to see that $\delta_S = 0$ may not hold. Instead, the valuations for the ambiguous acts, f and g, may both lie below that of the known mixture, $\frac{1}{2}h + \frac{1}{2}l$ such that $\delta_S < 0$. Consider the convex set of probability measures $P = [0,1]$. The minimal measure for act f has $\tilde{p}_{green} = 1$, yielding $U_{MEU}(f) = u(x)$. The minimal measure for act g is $\tilde{p}_{green} = 0$, yielding $U_{MEU}(g) = u(x)$. Hence, $\delta_S = \frac{1}{2}[u(x)+u(x)] - [\frac{1}{2}u(x)+\frac{1}{2}u(y)] < 0$.[14]

Second, for Subjective-Objective Consistency we recall the relaxed independence axiom, Certainty Independence, used in the MEU formulation. Certainty Independence states that independence holds for mixing subjective acts with objective acts. While MEU allows decision makers to entertain different beliefs when considering complementary subjective acts, it requires that unique beliefs be used to evaluate mixtures of a single subjective act with objective acts. For example, the MEU valuation of $\alpha f + (1-\alpha)h$ is

$$U_{MEU}(\alpha f+(1-\alpha)h) = \min_{\tilde{p}_{green}\in P}\left[\alpha(\tilde{p}_{green}\cdot u(y) + (1-\tilde{p}_{green})\cdot u(x)) + (1-\alpha)u(y)\right],$$

which is linear in α. Importantly, differences in valuations are, again, independent of subjective probabilities, e.g., $U_{MEU}(\alpha f + (1-\alpha)h) - U_{MEU}(\alpha f + (1-\alpha)l) =$

[14] The pattern $U_{MEU}(f), U_{MEU}(g) < U_{MEU}(\frac{1}{2}h + \frac{1}{2}l)$ is the only violation of Subjective-Subjective Consistency MEU allows. Since $\min_{p(s)\in P} p_s + \min_{p(s)\in P}(1-p_s) \leq 1$ it does not allow ambiguity seeking. This was one of the reasons for the development of its generalization, α-maxmin (Ghirardato et al., 2004), in which expected utility is calculated with respect to a convex combination of the worst-case and the best-case probability measure. The consistency violations implied by MEU will be shared by the more general α-MEU model.

$1 - U_{MEU}(\alpha h + (1-\alpha) l)$. For a fixed subjective act, an MEU agent's preferences over mixtures with objective acts will be indistinguishable from those of an SEU agent. Hence, Subjective-Objective Consistency will hold for MEU such that $\delta_M(\alpha, f) = 0$ and $\delta_M(\alpha, g) = 0$ for all α.

A model with similar consistency implications in our environment to MEU is Choquet Expected Utility (CEU, Schmeidler, 1989). Like MEU, CEU departs from Anscombe-Aumann by weakening the Independence axiom but does so in a slightly different way. Instead of demanding invariance to mixing with all acts (Independence) or objective acts (Certainty Independence), the theory's central axiom, termed Comonotonic Independence, demands invariance to mixtures with comonotonic acts only.[15] This yields a representation of the expected utility form in which subjective probabilities over events (termed *capacities*) are no longer additive.[16] Capacities for complementary events, like f and g, that add to less than or more than one deliver Subjective-Subjective inconsistencies.[17] For Subjective-Objective Consistency, we note that constant acts (including our simple objective acts) are comonotonic with all acts, hence Comonotonic Independence implies Certainty Independence in the presence of the other axioms (see Gilboa & Marinacci, 2011). Consequently, like MEU, CEU delivers consistency for subjective-objective mixtures.

Klibanoff et al. (2005) Smooth Ambiguity

Klibanoff et al. (2005, KMM for short) features a decision maker who has a proper probability distribution over states but does not treat the expected utilities obtained in each state equally. The utility functional is of the form

$$U_{KMM}(f) = \int_S p(s)\ \phi\left(u(a(s))\right)\ ds,$$

where $u(a(s))$ is the expected utility in state s and these expected utilities are then aggregated over states by the utility aggregation function $\phi(\cdot)$. First, for

[15] Two acts $a, b \in F$ are comonotonic if it is never the case that both $a(s) \succ a(s')$ and $b(s) \prec b(s')$ for some states of the world s and s'

[16] I.e. $p(A \cup B) \neq p(A) + p(B)$ for mutually exclusive events A and B

[17] Indeed, when capacities sum to one, the SEU form is recovered.

Subjective-Subjective Consistency, the model generates the inconsistencies of ambiguity aversion, $\delta_S < 0$, if $\phi(\cdot)$ is strictly concave, the model generates the inconsistencies of ambiguity seeking, $\delta_S > 0$, if $\phi(\cdot)$ is strictly convex, and the model collapses to SEU, $\delta_S = 0$, if $\phi(\cdot)$ is linear.[18]

For Subjective-Objective Consistency, KMM generates valuations for mixtures between subjective and objective acts that are not in general linear in mixture proportion.[19] Further, differences in mixture valuations will not generally be independent of subjective probabilities.[20] The shape of the aggregator function, $\phi(\cdot)$, determines whether $\delta_M(\alpha, f)$ and $\delta_M(\alpha, g)$ are greater than or less than

---

[18] To illustrate the principle with a concave $\phi(\cdot)$, assume for simplicity that the decision maker believes with probability $\frac{1}{2}$ that the ambiguous jar contains only green marbles and with probability $\frac{1}{2}$ that the urn contains only red marbles. The ambiguous acts f and g would then be evaluated as

$$\begin{aligned} U_{KMM}(f) &= \frac{1}{2}\phi\Big(1 \cdot u(y)\Big) + \frac{1}{2}\phi\Big(1 \cdot u(x)\Big) \\ = U_{KMM}(g) &= \frac{1}{2}\phi\Big(1 \cdot u(x)\Big) + \frac{1}{2}\phi\Big(1 \cdot u(y)\Big) \\ \leq U_{KMM}(\frac{1}{2}h + \frac{1}{2}l) &= \phi\left(\frac{1}{2} \cdot u(y) + \frac{1}{2} \cdot u(x)\right) \end{aligned}$$

where the inequality is a straight-forward consequence of the concavity of $\phi(\cdot)$.

[19] For simplicity, assume that the decision maker believes with probability $\tilde{p}$ that the unknown jar contains only green marbles and with probability $1 - \tilde{p}$ that the jar contains only red marbles. The act f would then be evaluated as

$$U_{KMM}(f) = \tilde{p}\phi\Big(1 \cdot u(y)\Big) + (1 - \tilde{p})\phi\Big(1 \cdot u(x)\Big)$$

Then mixtures involving f and h, for example, are

$$U_{KMM}(\alpha f + (1-\alpha)h) = \tilde{p}\phi\left(\alpha \cdot u(y) + (1-\alpha)u(y)\right) + (1-\tilde{p})\phi\left(\alpha \cdot u(x) + (1-\alpha)u(y)\right)$$

This form is linear in α only if $\phi(\cdot)$ is linear, i.e. if the KMM agent is SEU.

[20] Continuing the example where the decision maker believes with probability $\tilde{p}$ that the ambiguous urn contains only green balls and with probability $1 - \tilde{p}$ that the urn contains only red balls.

$$U_{KMM}(\alpha f + (1-\alpha)h) = \tilde{p}\phi\left(\alpha \cdot u(y) + (1-\alpha)u(y)\right) + (1-\tilde{p})\phi\left(\alpha \cdot u(x) + (1-\alpha)u(y)\right)$$

and

$$U_{KMM}(\alpha f + (1-\alpha)l) = \tilde{p}\phi\left(\alpha \cdot u(y) + (1-\alpha)u(x)\right) + (1-\tilde{p})\phi\left(\alpha \cdot u(x) + (1-\alpha)u(x)\right).$$

Table 3.1: Consistency Tests and Predictions

Consistency Test	Model			
	SEU	MEU	KMM	
			$\phi(\cdot)$ Concave	$\phi(\cdot)$ Convex
Subjective-Subjective Consistency:	Yes $\delta_S = 0$	No $\delta_S < 0$	No $\delta_S < 0$	No $\delta_S > 0$
Subjective-Objective Consistency:	Yes $\delta_M(\alpha, f) = 0$ $\delta_M(\alpha, g) = 0$	Yes $\delta_M(\alpha, f) = 0$ $\delta_M(\alpha, g) = 0$	No $\delta_M(\alpha, f) > 0$ $\delta_M(\alpha, g) > 0$	No $\delta_M(\alpha, f) < 0$ $\delta_M(\alpha, g) < 0$

zero. A concave (convex) $\phi(\cdot)$ function yields $\delta_M(\alpha, f)$, $\delta_M(\alpha, g)$ greater (less) than zero.[21] Note that the shape of $\phi(\cdot)$ determines the nature of inconsistencies for both Subjective-Subjective and Subjective-Objective Consistency. If $\phi(\cdot)$ is concave, one predicts, $\delta_S < 0$ and $\delta_M(\alpha, f), \delta_M(\alpha, g) > 0$, while if $\phi(\cdot)$ is convex, one predicts, $\delta_S > 0$ and $\delta_M(\alpha, f), \delta_M(\alpha, g) < 0$.

Table 3.1 summarizes our developments from SEU, MEU, and KMM models, providing both our definitions of consistency and evaluating the extent to which each is respected by each model. Note that non-SEU models are potentially separable by investigating both Subjective-Subjective and Subjective-Objective Consistency. While both MEU and KMM can rationalize ambiguity aversion, and so Subjective-Subjective inconsistencies, MEU maintains consistency for subjective objective mixtures. Epstein and Schneider (2010) provide an intuitive explanation for this distinction between KMM and other non-SEU models. The main difference between MEU and KMM is the kinds of mixtures that are valued by the decision maker. MEU holds that only mixtures between complementary subjective acts are

Hence,

$$\begin{aligned} U_{KMM}(\alpha f + (1-\alpha)h) - U_{KMM}(\alpha f + (1-\alpha)l) = \\ \tilde{p}\phi\left(\alpha \cdot u(y) + (1-\alpha)u(y)\right) + (1-\tilde{p})\phi\left(\alpha \cdot u(x) + (1-\alpha)u(y)\right) - \\ \tilde{p}\phi\left(\alpha \cdot u(y) + (1-\alpha)u(x)\right) + (1-\tilde{p})\phi\left(\alpha \cdot u(x) + (1-\alpha)u(x)\right). \end{aligned}$$

Note that the beliefs, $\tilde{p}$, do not cancel unless $\phi(\cdot)$ is linear.

[21] The proof, primarily graphical, is provided in Appendix 3.A1.

valuable because only such mixtures reduce the dependence of payoffs on the state of the world.[22] In contrast, in KMM mixing between a subjective act and another act can be valuable even if that act is objective. A KMM decision maker, in fact, even values mixing a subjective act with its certainty equivalent.

The objective of our experimental study is to develop empirical analogues for δ_S, $\delta_M(\alpha, f)$, and $\delta_M(\alpha, g)$. The next section describes our experimental environment and the operationalization of our consistency tests.

3.3 Experimental Design and Procedures

We experimentally test the consistency implications of Section 3.2 using a multiple price list style design.[23] Subjects make a series of binary choices between a potentially subjective act, say f, yielding outcomes x and $y > x$, and a known objective lottery yielding outcome y with probability q and outcome 0 with probability $1-q$. In each task, the probability q is varied from 0 to 1 in steps of 0.05. Hence, where an individual switches from preferring the given act, f, to the objective lottery yields interval information on the indifference condition

$$f \sim q_f \cdot h + (1 - q_f) \cdot 0,$$

and hence the valuation of the act.[24]

The decisions were introduced as a choice between drawing a marble from 'Jar A' and drawing a marble from 'Jar B'. Subjects were told *"For each decision,*

[22] As in the famous suggestion by Raiffa (1961) to simply flip a coin to decide which color from the Ellsberg urn to bet on, a strategy which reduces all subjective ambiguity to objective uncertainty

[23] The closest designs to our own are found in decision-making under objective uncertainty and were initially suggested in Farquhar's (1984) survey of utility assessment methods. They were implemented experimentally in one study of nine subjects using hypothetical monetary rewards (McCord & de Neufville, 1986), a number of medical questionnaires (Bleichrodt, Abellan-Perinan, Pinto-Prades, & Mendez-Martinez, 2007; Magat, Viscusi, & Huber, 1996; Oliver, 2005, 2007), and one incentivized experiment of decision-making under objective uncertainty. For decision-making with subjective events Baillon and Bleichrodt (2011) implements a similar method though with only two outcomes and refers to the task as "matching probabilities". Such methods also have close relations to the binary lottery procedure that is used in the literature on belief elicitation.

[24] A higher value of q_f corresponds to a higher valuation.

your task is to decide whether you prefer to draw a marble from JAR A or draw a marble from JAR B."[25] The contents of Jar A were potentially subjective while the contents of Jar B were always fully objective. Hence, where an individual switched from preferring to draw a marble from Jar A to preferring to draw from Jar B, carries interval information on q in the indifference condition above.

Specific colors linked exclusively to the contents of each jar. Red, green and white linked to the potentially subjective Jar A. Black and yellow linked to the objective Jar B, with black marbles yielding the high outcome, y, and yellow marbles yielding 0. Throughout the experiment the values $x = \$10$ and $y = \$30$ were maintained and each jar contained exactly 20 marbles. Figure 3.1 provides a sample task eliciting the valuation for the act $\frac{1}{2}g + \frac{1}{2}l$.

Eliciting valuations with lotteries in this manner has one key advantage over preference rankings or certainty equivalents. Under SEU, the equivalent lottery for f, q_f, is a linear function of the subjective expected utility of the act. That is, the equivalent lotteries preserve not just the ordinal ranking of expected utilities but also its cardinal scale. Hence, when $u(0)$ and $u(y)$ are normalized to 0 and 1 respectively, $q_f = U_{SEU}(f)$. This stands in contrast to certainty equivalents which are non-linear functions of expected utility for any non-linear utility function. Under SEU we can directly use the elicited equivalent probabilities to calculate subjective premia and mixture differences. We denote these empirical analogues to our consistency measures by d_S, $d_M(\alpha, f)$ and $d_M(\alpha, g)$.[26]

For MEU, under the normalization above, it is also the case that $q_f = U_{MEU}(f)$, such that valuations translate directly to consistency measures. However, under KMM the relationship between utility and equivalent probabilities is confounded by the shape of the utility aggregator, $\phi(\cdot)$. Under the same normalization as above $U_{KMM}(f) = \phi(q_f)$ and so $q_f = \phi^{-1}(U_{KMM}(f))$. Though the implications

[25] The full experimental instructions are provided in Appendix S.3.1.

[26] Specifically,

$$d_S = \frac{1}{2}[q_g + q_f] - q_{\frac{1}{2}h+\frac{1}{2}l},$$

$$d_M(\alpha, g) = [q_{\alpha g+(1-\alpha)h} - q_{\alpha g+(1-\alpha)l}] - [1 - q_{\alpha h+(1-\alpha)l}],$$

and

$$d_M(\alpha, f) = [q_{\alpha f+(1-\alpha)h} - q_{\alpha f+(1-\alpha)l}] - [1 - q_{\alpha h+(1-\alpha)l}].$$

Note that this assumes $q_h = 1$.

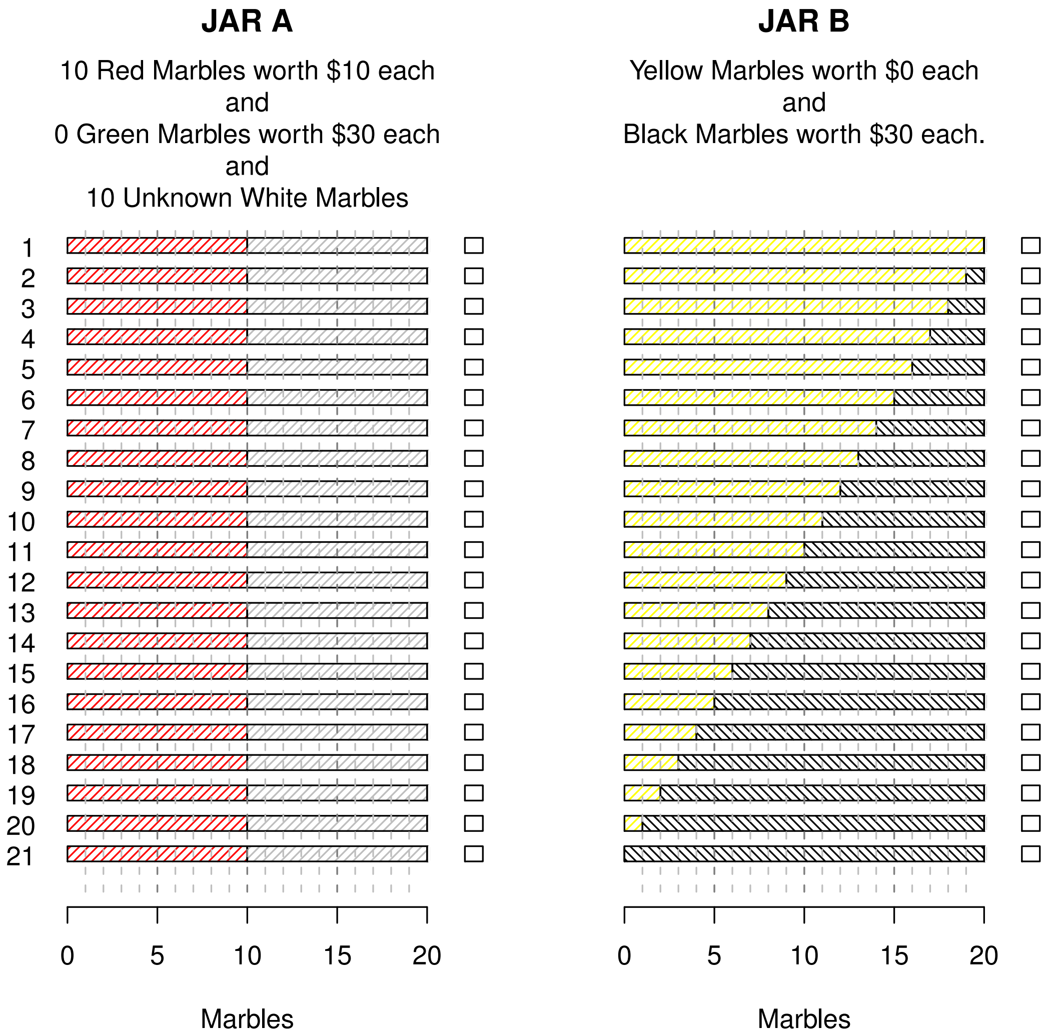


Figure 3.1: Multiple price list for task 4: $\frac{1}{2}g + \frac{1}{2}l$

for δ_S are maintained, since $U_{KMM}(f)$ is a concave function of α and $\phi^{-1}(\cdot)$ is a convex function of its argument, the full set of implications for the mixture differences $\delta_M(\alpha, f)$ and $\delta_M(\alpha, g)$ displayed in Table 3.1 do not carry through to the empirical analogues of these measures, $d_M(\alpha, f)$ and $d_M(\alpha, g)$. As is shown in Appendix 3.A2, the sign of these empirical measures turns on the absolute concavity of the utility aggregator, $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$. Two functional forms for $\phi(\cdot)$, prominent in the literature on smooth ambiguity aversion are exponential, $\phi(x) = -e^{-\theta x}$, and power, $\phi(x) = \frac{x^{1-\theta}}{1-\theta}$. For the exponential family, $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$ is constant in its argument and mixture distances will be zero for all α. That is, in our experiment, exponential KMM and MEU are not separable. For the power family, when $\theta > 0$ and thus $d_S < 0$, $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$ is decreasing, and empirical mixture distances, $d_M(\alpha, f)$ and $d_M(\alpha, g)$, will be positive. Negative mixture distances can be delivered only if $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$ is increasing in its argument.

Another important point is that the absolute concavity of the aggregator, $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$, is distinct from the sign of the second derivative, $\phi''(\cdot)$. As such, for KMM the empirical predictions with respect to the mixture distances, $d_M(\alpha, f)$ and $d_M(\alpha, g)$, are distinct from the predictions for d_S, which turns only on the sign of $\phi''(\cdot)$. Stated otherwise, KMM is flexible enough in our environment to rationalize any pattern of behavior given choice of the aggregator. Thus, our exercise with respect to KMM will be reduced to analyzing the consistency of the data with the popularized functional forms for the aggregator function: exponential and power.

3.3.1 Implementing Ambiguity

In order to identify valuations for subjective acts and mixtures of subjective and objective acts, an experimental device was introduced to implement ambiguity. Prior to making any experimental decisions, subjects were told that the colors of some marbles of Jar A may be unknown to them. These marbles of unknown color would be represented by white marbles. This is the case, for example, with act $\frac{1}{2}g + \frac{1}{2}l$ represented in Figure 3.1. Hence, Jar A could be described as having some red marbles, some green marbles and some unknown white marbles that could be red or green.

Because valuations for multiple subjective acts and subjective-objective mixtures were elicited, we found it critical to attempt to fix beliefs for the unknown white marbles. Hence, subjects were introduced to a third jar, Jar X. Jar X was in the front of the laboratory at the beginning of the experiment and was not moved throughout. Jar X was surrounded by a cylindrical piece of construction paper, obscuring its contents. Subjects were told

"There is a third jar on the table at the front of the room right now,

> *marked JAR X. JAR X contains 20 marbles. These marbles are some combination of red and green and have already been determined. There may be anywhere from 0 red marbles and 20 green marbles to 20 red marbles and 0 green marbles in JAR X, or any combination of 20 red and green marbles."*[27]

If an unknown white marble was drawn from Jar A it would induce a draw from Jar X to determine its color and hence payment.[28] Subjects were then told *"A number of questions will refer to the unknown white marbles and so refer to JAR X. You should think carefully about the possible contents of JAR X."*

The experimental device for implementing ambiguity allows for the elicitation of valuations for acts f and g by varying only the labeling of payment for Jar A. That is, for act f a red marble pays \$30 and a green marble pays \$10, while for act g a red marble pays \$10 and a green marble pays \$30. For each of these purely subjective acts, Jar A consists only of white marbles, hence one draws from the unknown Jar X with probability 1. Given that Jar X is unchanged between the two elicitations, beliefs are plausibly fixed and behavior should accord with SEU.[29]

The experimental device also allows for subjective-objective mixtures by leaving payment labels unchanged and varying only the contents of Jar A. For example to implement the act $\frac{3}{4}g + \frac{1}{4}h$, Jar A would consist of 15 unknown white marbles and 5 green marbles, with green marbles yielding \$30 and red marbles yielding \$10.[30]

[27] Incidentally, the contents of Jar X for each session were determined by a visiting, overqualified research assistant, expert in decision theory, DD.

[28] Specifically subjects were told *"If a white marble is drawn from JAR A, we will draw a marble from JAR X. If the marble drawn from JAR X is green you will receive the payment for a green marble ... If the marble drawn from JAR X is red you will receive the payment for a red marble. "*

[29] Offering subjects bets with both red and green as the winning color should also dispel any suspicion amongst subjects that the composition of Jar X was somehow rigged by the experimenters, which has been identified in several recent studies as a potential confound (Binmore, Stewart, & Alex Voorhoeve, 2013; Charness, Karni, & Levin, 2012).

[30] This implementation was inspired by the concept of Certainty Independence as the subject's exposure to the subjective act (the unknown Jar X) is reduced by the construction of Jar A. The design differs notably from that used by Hong et al. (2013) to construct "partial ambiguity". In these experiments subjects were presented with a decks of cards with varying compositions. Subjects were told that the number of winning cards would either lie within a certain range (e.g. between 60 and 80 cards out of 100) or be one of two possibilities (e.g. either 40 out of 100 or 60 out of 100). Since subjects were presented with an entirely different deck of cards in every round beliefs about the subjective component of the bets may not have been fixed.

Note that with the exception of fully subjective and fully objective acts, a draw from Jar A induces a compound gamble. Importantly, for the theories described in Section 3.2 all predictions are maintained as gambles are appropriately compounded at the stage in question.

3.3.2 Design Details

In total, subjects faced 20 experimental tasks separated into three blocks. In the first block, red marbles yielded \$10 and green marbles yielded \$30. Eight tasks elicited the valuation for subjective act g, and mixtures of g with h and l. The proportion, α, of act g in each mixture varied from 1 to 0.25. In one task, task 5, the objective portion of the mixture was not degenerate, but rather $\frac{1}{2}l + \frac{1}{2}h$. Table 3.2, Panel A provides the tasks.

The third block mirrored the first, except red marbles yielded \$30 and green marbles yielded \$10. Eight tasks elicited the valuation for subjective act f, and mixtures of f with h and l. The proportion, α, of act f in each mixture varied from 1 to 0.25. In one task, task 17, the objective portion of the mixture was not degenerate, but rather $\frac{1}{2}l + \frac{1}{2}h$. Table 3.2, Panel C provides the tasks.

In the second block, valuations of purely objective acts are elicited. That is, $\alpha = 0$ and the proportion of valuations for the acts $\frac{1}{4}l + \frac{3}{4}h$, $\frac{1}{2}l + \frac{1}{2}h$, $\frac{3}{4}l + \frac{1}{4}h$, and $\frac{1}{1}l + \frac{0}{1}h$ were elicited. Table 3.2, Panel B provides the tasks.

Two orders of the blocks were implemented at the session level to identify potential order effects: Block 1, Block 2, Block 3; and Block 3, Block 2, Block 1. The objective acts of Block 2 were maintained as a buffer between the two similar subjective act task blocks.[31] Each block was distributed as a packet of decision sheets and were collected upon completion. Once all subjects had completed a block, the next was distributed. Within a block, subjects could work in any order they chose and each task had specific instructions written at the top of the task sheet.[32]

Six sessions with a total of 133 subjects were conducted at the University of California, San Diego in October of 2011. The experiment was pencil and paper

[31] No order or session effects were observed.
[32] See Appendix S.3.1 for the full set of instructions.

Table 3.2: Experimental Decisions

Panel A (Block 1)		Panel B (Block 2)		Panel C (Block 3)	
Subjective Act g (Red, Green) pays (10, 30)		Objective Acts l and h (Red, Green) pays (10, 30)		Subjective Act f (Red, Green) pays (30, 10)	
Task Number	Act Description	Task Number	Act Description	Task Number	Act Description
1	$\frac{1}{4}g + \frac{3}{4}h$	9	$\frac{1}{4}l + \frac{3}{4}h$	13	$\frac{1}{4}f + \frac{3}{4}h$
2	$\frac{1}{4}g + \frac{3}{4}l$	10	$\frac{1}{2}l + \frac{1}{2}h$	14	$\frac{1}{4}f + \frac{3}{4}l$
3	$\frac{1}{2}g + \frac{1}{2}h$	11	$\frac{3}{4}l + \frac{1}{4}h$	15	$\frac{1}{2}f + \frac{1}{2}h$
4	$\frac{1}{2}g + \frac{1}{2}l$	12	l	16	$\frac{1}{2}f + \frac{1}{2}l$
5	$\frac{1}{2}g + \frac{1}{2}(\frac{1}{2}h + \frac{1}{2}l)$			17	$\frac{1}{2}f + \frac{1}{2}(\frac{1}{2}l + \frac{1}{2}h)$
6	$\frac{3}{4}g + \frac{1}{4}l$			18	$\frac{3}{4}f + \frac{1}{4}l$
7	$\frac{3}{4}g + \frac{1}{4}h$			19	$\frac{3}{4}f + \frac{1}{4}h$
8	g			20	f

based and subjects were provided with calculators to aid them with any calculations. In addition to any experimental earnings, subjects received a \$5 minimum payment. Including minimum payments, average subject earnings were \$29.89 and the experimental duration was around 90 minutes.

3.3.3 Implementing Payment

A random incentive mechanism was used to implement payment. One decision from one task was chosen to be the ‘decision-that-counts’. With 21 decisions per task and 20 tasks, each decision had a 1 in 420 chance of being the decision-that-counts. Subjects were told in advance the following order of implementation: the decision-that-counts would be chosen, then the corresponding Jar A and Jar B would be constructed, then a marble would be drawn from the Jar corresponding to choice.[33]

This mechanism is commonly used in the experimental literature on ambiguity aversion (see e.g. Abdellaoui, Klibanoff, & Placido, 2011; Binmore et al., 2013; Charness et al., 2012). Nonetheless there is some question as to whether the mechanism satisfies isolation, i.e. whether choices made under the mechanism are identical to choices that would be made if subjects were presented with only a single choice.

Neither the theoretical nor the empirical literature on the issue have so far provided a definitive answer to the question. The theoretical literature on choice under risk, dating to Holt (1986) and Karni and Safra (1987), has suggested that random mechanisms need not be incentive compatible for choices between risky prospects if either the Independence or Reduction of Compound Lotteries axioms are violated. If ambiguity aversion is caused by violations of Independence one may therefore suspect that random mechanisms cannot possibly elicit ambiguity averse preferences. Indeed, Bade (2014) proves an impossibility result for the elicitation of ambiguity preferences under a random incentive mechanism. Fortunately, things may not be quite as dire. Both Azriely, Chambers, and Healy (2014) and Baillon, Halevy, and Li (2014) show that the random incentive mechanism can be incentive compatible even if subjects are offered bets on complementary acts.

The key issue in this theoretical debate is whether the random payment mechanism affords a subject the opportunity to hedge against ambiguity. For example, consider only two tasks eliciting the valuations only for the complementary acts f and g. This presents a case in which the logic in Raiffa (1961) suggests subjects should perfectly hedge via the mechanism. A subject could, in principle, state the highest possible valuation in each case, choosing the subjective act in every binary decision. Then the random payment mechanism would give this subject either act f or act g with equal probability. Regardless of the color of the drawn marble the individual is holding an objective 50-50 gamble, hedging completely. Two things are of note. First, such opportunities should make subjective acts seemingly more attractive, leading to ambiguity seeking behavior. Hence, the extent to which

[33] In practice one ‘decision-that-counts’ was chosen for each session, the jars were constructed and a ball was drawn from each to determine the payments for individuals that had preferred each. Hence, in practice subjects did know the realization from their unchosen jar in the decision-that-counts.

the data exhibit ambiguity aversion may be instructive on the prevalence of such strategies. Second, there is a nuance in the argument above with respect to the order of conditioning (Azriely et al., 2014; Baillon et al., 2014). An individual is holding an objective 50-50 gamble only if the task implementing payment is chosen after the state (drawn marble color) is realized. If the task implementing payment is chosen first, then the opportunity to hedge is plausibly reduced as for a given task an individual exploiting the strategy above continues to hold a subjective bet. Note that our order of implementation chooses the decision-that-counts first, then constructs the corresponding jars, and then draws marbles, potentially limiting such hedging opportunities. Reassuringly, in an application that is very close to ours, Binmore et al. (2013) find no evidence that choices are affected by the opportunity to hedge via the mechanism.

3.4 Results

The results are presented in two sub-sections. First, we investigate aggregate decision-making describing the extent to which average behavior adheres to the requirements outlined in Section 3.2 for Subjective-Subjective Consistency and Subjective-Objective Consistency. Second, we investigate individual consistency measures and explore the extent of correlation between Subjective-Subjective and Subjective-Objective inconsistencies. We also document a further deviation from SEU in our environment: the presence of inconsistent valuations between subjective and objective acts with a dominance relation. In order to describe our inconsistencies via the beliefs and utility parameters that rationalize behavior, throughout we make appeal to an exploratory appendix (Appendix 3.A4), which estimates decision-making parameters based on the observed stylized facts.

Of 133 subjects who participated in the experiment, 22 (17%) subjects exhibited multiple switching points in at least one task.[34] Such multiple switching is frequently found in multiple price list experiments and normally occurs for 10-15% of subjects (Holt & Laury, 2002; Meier & Sprenger, 2010). Our analysis focuses on the 111 subjects who completed all aspects of the experiment without multiple

[34] Of these 22, 13 subjects failed to indicate a unique switching point exactly once. There does not seem to be a systematic pattern to the tasks for which such multiple switches occurred.

switching. However, the results are virtually unchanged when including all 133 subjects and removing only the errant data.

3.4.1 Aggregate Results

Figure 3.2 presents the aggregate data graphically, providing the mean valuation, $q \times 100$, and corresponding standard error for each task. These valuations are expressed in probability units of \$30 such that one can say the mean subject would be indifferent between act g and a 69.8% chance of receiving \$30. Each panel corresponds to a distinct block in the experiment and each series corresponds to a mixture. For example, Panel A corresponds to Block 1 of the experiment, eliciting valuations for act g and mixtures with h and l. In act g a green marble pays \$30 and a red marble pays \$10. The red series links mixtures $\alpha g+(1-\alpha)l$, corresponding to mixing g with low-paying red marbles. The green series links mixtures $\alpha g+(1-\alpha)h$, corresponding to mixing g with high-paying red marbles. The mixture proportion, α, is on the horizontal axis and the corresponding valuations are on the vertical axis. The two series come together at $\alpha = 1$ corresponding to the fully subjective act, g. Panel C provides the same information corresponding to Block 3 of the experiment, but with the labeling reversed to reflect that for act f red marbles pay \$30 and green marbles pay \$10. Panel B provides the data for Block 2 graphing $\lambda l + (1 - \lambda)h$ against their valuations. The proportion of low-paying balls, λ, is on the x-axis.

The multiple price lists we use to elicit valuations only identify the point at which a subject's preference switches from Jar A to Jar B to within a 5% interval. To account for this interval censoring of the data, we estimate an interval regression with indicator variables for each task (Stewart, 1983) and standard errors clustered at the individual level to capture arbitrary correlation in errors within individual across tasks. The corresponding mean and standard error are calculated for each task, generating the point and standard error bars presented in Figure 3.2.[35] We use the estimates from similar interval regressions to statistically test our consistency requirements.

Several features of Figure 3.2 are worthy of attention. First, note that a coherent

[35] This regression table is reproduced as Appendix Table 3.A1.

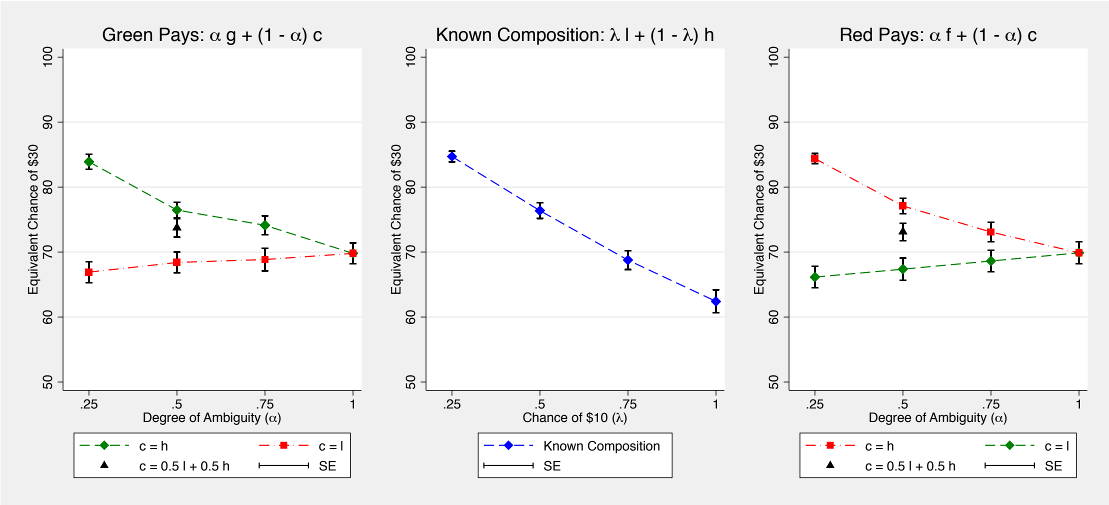


Figure 3.2: Estimated mean valuations and standard errors from interval regressions (Stewart, 1983) for all experimental choices. See Appendix Table 3.A3 for further detail.

pattern of mixtures appears in Panels A and C. Mixtures of a subjective act with h yield increased valuations while mixtures with l yield decreased valuations. Second, though four points in each series are likely too few to make conclusive statements, each series is approximately linear in the object of interest, either α or λ. Though such data suggest comprehension and systematic response for each series, it is the links between the series that deliver our consistency tests.

Subjective-Subjective Consistency Our test for Subjective-Subjective Consistency compares the average valuation of f and g to the valuation of $\frac{1}{2}h + \frac{1}{2}l$. The mean valuation of act f is $q_f = 69.89$ (clustered s.e. $= 1.70$) and the mean valuation of act g is $q_g = 69.80$ (1.60). The mean valuation of $\frac{1}{2}h + \frac{1}{2}l$ is $q_{\frac{1}{2}h+\frac{1}{2}l} = 76.37$ (1.20). We construct the emprical average subjective premium, $d_S = \frac{1}{2}[q_f + q_g] - q_{\frac{1}{2}h+\frac{1}{2}l} = -6.53$ (1.13). Under SEU this represents the subjective expected utilities of each gamble such that d_S is predicted to be zero. We reject the null hypothesis of zero at all conventional levels ($\chi^2(1) = 33.4$, $p < 0.01$).[36]

We find d_S of around 6.5 percentage points, or around 9.4% of the average valuation of the subjective acts in question. Valuing the average of subjective acts lower

[36] All hypotheses tests are reported in Appendix Table 3.A1.

than their objective counterpart is evidence of $\delta_S < 0$, or ambiguity aversion. As in other Ellsberg style experiments, regardless of the labeling of the states, subjective acts yield lower valuations than an objective 50-50 mixture. In Appendix 3.A4 we provide estimates of utility parameters. In our preferred specification, Table 3.A2, column (4), belief estimates indicate that act f is believed to be 69.2% (clustered s.e = 3.5%) low-paying green marbles while act g is believed to be 69.5% (3.7%) low-paying red marbles. The null hypothesis of additive beliefs under subjective expected utility is rejected at all conventional levels, ($\chi^2(1) = 41.1,\ p < 0.01$).

Subjective-Objective Consistency We turn next to Subjective-Objective Consistency examining mixtures between subjective and objective acts. Consider Panel A of Figure 3.2 with $\alpha = 0.5$, which presents a convenient example. The mean valuation for $\frac{1}{2}g+\frac{1}{2}h$ is $q_{\frac{1}{2}g+\frac{1}{2}h} = 76.46$ (1.21) while the mean valuation of $\frac{1}{2}g+\frac{1}{2}l$ is $q_{\frac{1}{2}g+\frac{1}{2}l} = 68.40$ (1.61). These values are interestingly compared to the mean valuations of $\frac{1}{2}h + \frac{1}{2}l$, $q_{\frac{1}{2}h+\frac{1}{2}l} = 76.37$ (1.20), and l, $q_l = 62.41$ (1.76). Such comparisons already demonstrate potential inconsistencies in the treatment of the subjective act, g, depending on whether it is mixed with a high or a low outcome.

To begin, note that the valuations of $\frac{1}{2}g + \frac{1}{2}h$ and $\frac{1}{2}h + \frac{1}{2}l$ are almost identical. Such similarity would occur if subjects believed that g delivered only low-paying marbles.[37] However if g delivered only low-paying marbles then the valuations of $\frac{1}{2}g + \frac{1}{2}l$ and l should be identical. Instead $q_{\frac{1}{2}g+\frac{1}{2}l}$ lies significantly above q_l, ($\chi^2(1) = 21.87,\ p < 0.01$). The closest purely objective valuation to $q_{\frac{1}{2}g+\frac{1}{2}l}$ is for $\frac{1}{4}h + \frac{3}{4}l$, $q_{\frac{1}{4}h+\frac{3}{4}l} = 68.76(1.45)$. Such similarity would occur if subjects believed that g delivered 50% low-paying marbles and 50% high-paying marbles.[38] This initial comparison provides a compelling suggestion of Subjective-Objective inconsistency. Individuals appear to exhibit a “directed pessimism", exhibiting more pessimism when mixing with a high outcome than when mixing with a low outcome.

To explore further, for every α we calculate the average $d_M(\alpha, f)$ and $d_M(\alpha, g)$ from linear combinations of interval regression coefficients with standard errors

[37] Hence, the act $\frac{1}{2}g + \frac{1}{2}h$ would deliver $\frac{1}{2}$ high-paying marbles and $\frac{1}{2}$ low-paying marbles.
[38] Hence, the act $\frac{1}{2}g + \frac{1}{2}l$ would deliver $\frac{1}{4}$ high-paying marbles and $\frac{3}{4}$ low-paying marbles.

Table 3.3: Estimated Mixture Distances

		α		
		0.25	0.5	0.75
Act f	$d_M(\alpha, f)$	-13.00***	-13.90***	-10.83***
		(1.10)	(1.26)	(1.17)
Act g	$d_M(\alpha, g)$	-14.26***	-15.56***	-10.02***
		(1.36)	(1.16)	(1.06)

Average distance measures calculated from linear combinations of regression coefficients after interval regression with standard errors clustered at individual level. Regression table available as Appendix Table 3.A1. All values tested for $H_0 := 0$. *Level of significance*: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

clustered at the individual level.[39] The average values and standard errors are reported in Table 3.3. Table 3.3 reveals a clear pattern of inconsistency. For every act and every mixture, we reject the null hypothesis that the mixture distances are zero. Depending on the act and mixture, the empirical mixture distance lies between 10 and 15 percentage points below the SEU prediction. As demonstrated above, such negative distances are consistent with subjects acting as if they place a significantly lower likelihood of the subjective act yielding a high-paying marble when mixing with h than when mixing with l. In Appendix 3.A4 we provide estimates of utility parameters. In our preferred specification, Table 3.A2, column (4), belief estimates are supportive of a directed pessimism. When mixing with h, subjective acts are believed to consist of around 90% low-paying marbles, while when mixing with l they are believed to consist of around 55% low-paying marbles. The null hypothesis of consistent mixing is rejected at all conventional levels, ($\chi^2(6) = 109.4,\ p < 0.01$).

The aggregate data are broadly inconsistent with SEU, but rather indicate both Subjective-Subjective and Subjective-Objective inconsistencies. These deviations from subjective expected utility are not necessarily well-accommodated by other models. Consider first MEU, which predicts $d_S < 0$ and $d_M(\alpha, g),\ d_M(\alpha, f) = 0$ for

[39] For example, the regression yields values for $q_{\frac{1}{2}g+\frac{1}{2}l}$, $q_{\frac{1}{2}g+\frac{1}{2}h}$, and $q_{\frac{1}{2}h+\frac{1}{2}l}$. The value $d_M(\alpha, g)$ is calculated as $q_{\frac{1}{2}g+\frac{1}{2}h} - q_{\frac{1}{2}g+\frac{1}{2}h} + q_{\frac{1}{2}h+\frac{1}{2}l} - 1$. The corresponding regression table is available as Appendix Table 3.A1 with all corresponding calculations.

all α. Though the data adhere with the first prediction, the second is not supported as we find negative mixture distances inconsistent with MEU preferences. Next, consider KMM preferences, which can accommodate both Subjective-Subjective and Subjective-Objective inconsistencies. However, the presence of both $d_S < 0$ and $d_M(\alpha, g),\ d_M(\alpha, f) > 0$ requires a concave aggregator, $\phi(\cdot)$, to rationalize $d_S < 0$, and increasing absolute concavity, $-\frac{\phi''(\cdot)}{\phi'(\cdot)}$, to rationalize $d_M(\alpha, g),\ d_M(\alpha, f) > 0$. Neither popularized functional form for KMM delivers such behavior. Hence, the observed effects are inconsistent with popular formulations for KMM preferences. Given the flexibility of the KMM model, however, one could view our results as potentially disciplining future KMM applications to functional forms satisfying the properties above.

We now turn to individual level analyses attempting to measure both the heterogeneity in violations across individuals and the linkages between inconsistency measures.

3.4.2 Individual Results

We document widespread violations of our two SEU consistency requirements in the aggregate. The present subsection analyzes individual heterogeneity in behavior, a topic of recent interest (see, e.g. Abdellaoui, Baillon, Placido, & Wakker, 2010; Halevy, 2007). We also explore the interrelations between inconsistencies to shed light on a potential unifying explanation for our results. Our individual results broadly reproduce the aggregate data showing evidence of both Subjective-Subjective and Subjective-Objective inconsistencies. The two phenomena, however, are largely independent. The individual analysis documents an additional feature of our data: the presence of inconsistent valuations between subjective and objective acts with a dominance relation.

Subjective-Subjective Consistency The aggregate data show an average valuation for an ambiguous urn to be around 10% below its objective counterpart. Though this helpfully demonstrates the strength of preference in the aggregate data, the mean behavior masks substantial heterogeneity. In order to analyze individual behavior, we calculate the individual level Subjective Premium $d_{S,i} =$

$\frac{1}{2}(q_{f,i} + q_{g,i}) - q_{\frac{1}{2}h+\frac{1}{2}l,i}$, based upon the midpoint of individual responses. Figure 3.3, Panel A presents the distribution of $d_{S,i}$. 23 of 111 (21%) have $d_{S,i} = 0$ while 20 (18%) have $d_{S,i} > 0$, indicating ambiguity seeking behavior. The remaining 68 subjects (61%) have $d_{S,i} < 0$, indicating ambiguity aversion.[40]

Subjective-Objective Consistency The aggregate data show evidence of Subjective-Objective inconsistency. Subjective bets are not treated the same way when mixed with a high versus a low outcome. In order to develop an individual measure of subjective-objective consistency, we calculate the average individual level distance over the six subjective-objective mixtures based upon the midpoint of their responses, $d_{M,i}$.[41] Figure 3.3, Panel B presents the distribution of $d_{M,i}$. 105 of 111 (94.6%) exhibit $d_{M,i} < 0$. The average distance is -12.92, again consistent with the aggregate behavior of a directed pessimism.

Interrelation: We document two patterns of behavior inconsistent with SEU and mutually inconsistent with popularized non-SEU theories. However, the data, at least in the aggregate appear to be systematically generated. Here we examine the extent of correlation between our patterns of behavior. Figure 3.3, Panel C presents the individual level correlation between $d_{S,i}$ and $d_{M,i}$. A negative and significant correlation is observed wherein those subjects who exhibit the patterns of ambiguity seeking, $d_{S,i} > 0$ have more negative distance measures, $\rho = -0.233,\ p < 0.05$. However, this raw correlation is largely driven by outliers. Insignificant correlations are found when conducting a 1% trim of either $d_{S,i}$ or $d_{M,i}$.[42] Appendix Table 3.A1 also provides the aggregate regressions separately for the groups $d_{S,i} > 0$, $d_{S,i} = 0$, and $d_{S,i} < 0$. Negative mixture distances are observed for all three groups though they appear somewhat more pronounced for

[40] Appendix Table 3.A1 also provides the aggregate regressions separately for these three groups. Individuals with $d_{S,i} < 0$ have an average strength of preference of 12.9 percentage points for the objective jar, while individuals with $d_{S,i} > 0$ have an average strength of preference of 7.5 percentage points for the subjective jar.

[41] That is for a given individual $d_{M,i} = \frac{1}{6}\sum_{\alpha}\sum_{j\in g,f} d_M(\alpha,j)_i$ and $d_M(\alpha,j)_i = q_{\alpha j+1-\alpha h,i} - q_{\alpha j+1-\alpha l,i} + q_{\alpha h+1-\alpha l,i} - 1$. Taking the individual mean is a natural measure as a high degree of intercorrelation is observed between the six individual mixture distances, Cronbach's $\alpha = 0.78$.

[42] With a 1% trim of $d_{S,i}$ the correlation with 105 subjects is $\rho = -0.14,\ p = 0.15$. With a 1% trim of $d_{M,i}$ the correlation with 107 subjects is $\rho = -0.13,\ p = 0.18$. With a 1% trim along both dimensions, the correlation with 103 subjects is $\rho = -0.12,\ p = 0.24$.

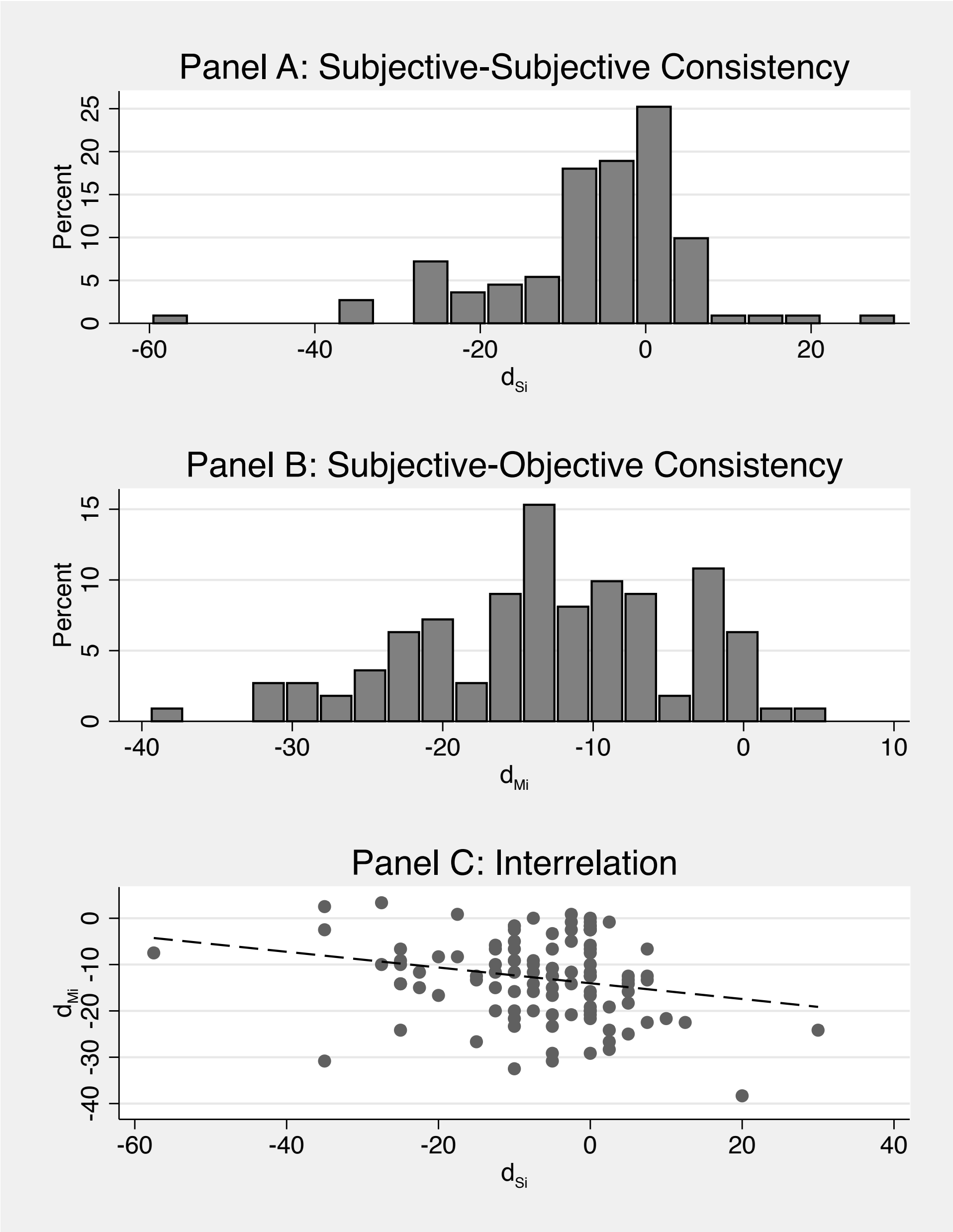


Figure 3.3: Individual Data

the 20 individuals with $d_{S,i} > 0$. We cannot reject the null hypothesis that those subjects with $d_{S,i} < 0$ and $d_{S,i} = 0$ have the same mixture distance values, but those subjects with $d_{S,i} > 0$ and $d_{S,i} < 0$ do differ significantly.

The patterns of and correlations in deviations we uncover are inconsistent with both MEU and KMM preferences. First, MEU predicts zero Subjective-Objective inconsistencies which is soundly rejected in the data. Second, KMM under the exponential formulation would predict the same pattern and so is also rejected in the data. KMM under the power formulation with $\theta > 0$ predicts that if $d_{S,i} < 0$, then $d_{M,i} > 0$, and there should exist a negative correlation between the two.[43] Though the full data set do provide a negative correlation, the correlation is not robust and the general location of the data ($d_{Mi} < 0$) is not in accord with prediction. An alternative formulation of KMM featuring a concave aggregator with increasing absolute concavity may be able to match broad features of the data, but the lack of correlation between $d_{S,i} < 0$ and $d_{M,i} > 0$ potentially suggests a formulation where these two features are independent.

Additional Inconsistencies: Dominance In addition to Subjective-Subjective and Subjective-Objective inconsistencies, the individual data afford the opportunity to investigate other potentially interesting patterns. Chief among these are situations in which one can compare a subjective-objective mixture with a purely objective act for which a dominance relation maintains. Indeed, our data have many such situations and we were surprised to see substantial violations of dominance.

We investigate situations in which individuals value the subjective portion of an act, g, higher than it's best possible outcome, h, or lower than its worst possible outcome, l. That is, we investigate whether the following inequalities are violated:

$$q_{\alpha l+(1-\alpha)h} < q_{\alpha g+(1-\alpha)h} < q_{\alpha h+(1-\alpha)h}.$$

Each one of these comparisons is indirect in the sense that the valuations come from different tasks.

To begin, 10 of 111 subjects (9%) have lower valuations for f and g than for l, valuing an ambiguous draw lower than its worst possible outcome.[44] In moving along the series of mixtures between given acts and h and l subjects have additional

[43] That is under the power formulation, the greater is θ, the lower is $d_{S,i}$ and the higher is $d_{M,i}$.
[44] Such individuals have disproportionately lower values for $d_{S,i}$ with a mean of -20.5 (s.d. =16.91) relative to the average of the other 101 subjects of -5.15 (10.43), ($t(109) = 4.17$, $p < 0.01$). No difference in mixture distances is observed.

opportunities to violate dominance either by valuing the subjective portion too low (in the case of h) or valuing it too high (in the case of l). In total, subjects have 6 opportunities to violate dominance when mixing with h and 6 opportunities when mixing with l. Interestingly, different subjects violate dominance in different ways and do so on multiple occasions.

Table 3.4 presents count measures of dominance violations and the relationship between dominance violations and $\delta_{S,i}$. Dominance violations are frequent throughout the data, and are substantially more prevalent when considering mixtures with h as oppposed to l.[45] When mixing with the high-paying act h violations are more prevalent among ambiguity averse, $\delta_{S,i} < 0$, subjects ($p < 0.01$, two-sample Kolmogorov-Smirnov test). Indeed, among $\delta_{S,i} < 0$ subjects only 16% never violate dominance while this is the case for more than 70% of the non-averse subjects. For mixtures involving the low-paying act l, in contrast, ambiguity seeking subjects $\delta_{S,i} > 0$ are more likely to violate dominance ($p < 0.01$, two-sample Kolmogorov-Smirnov test).[46]

An initial view of our dominance violations suggests subject errors, potentially indicative of subject confusion. Though this is one potential explanation, the extent of correlation with frequently observed experimental findings such as Ellsberg paradox behavior is a sign of regularity. Hence, we are hesitant to draw such conclusions. Instead, we recognize the inconsistency of dominance violations with all considered models, both SEU and non-SEU, and present the violations as evidence of an incomplete understanding of attitudes towards ambiguity which perhaps future work can rationalize.

Our dominance violations are observed when comparing subjective to objective bet evaluations. The majority of violations are in the direction of valuing a subjective act below its worst possible objective counterpart. For such observations, beliefs and risk aversion will not be sufficient to rationalize the data.[47] As these

[45] 80% of subjects have zero or one violation of the form where the subjective act is valued higher than its best possible outcome (mixing with l), while only 55% of subjects have zero or one violation of the form where the subjective act is valued lower than its worst possible outcome (mixing with h).

[46] There is also a significant correlation between dominance violations when mixing with l and $d_{M,i}$, ($\rho = -0.504,\ p < 0.01$). No significant correlation exists for dominance violations when mixing with h and $d_{M,i}$, ($\rho = -0.034,\ p = 0.72$).

[47] Violating dominance is akin to believing a subjective act as consisting of more than 100%

		0	1	2	3	4	5	6
Dominance Violations when mixing with h	$\delta_{S,i} > 0$(N=20)	70.00	15.00	10.00	5.00	0.00	0.00	0.00
	$\delta_{S,i} = 0$ (N=23)	78.26	13.04	8.70	0.00	0.00	0.00	0.00
	$\delta_{S,i} < 0$ (N=68)	16.18	19.12	19.12	23.53	14.71	2.94	4.41
	Total (N=111)	38.74	17.12	15.32	15.32	9.01	1.80	2.70
Dominance Violations when mixing with l	$\delta_{S,i} > 0$ (N=20)	10.00	25.00	25.00	20.00	10.00	5.00	5.00
	$\delta_{S,i} > 0$ (N=23)	73.91	13.04	4.35	4.35	0.00	4.35	0.00
	$\delta_{S,i} < 0$ (N=68)	69.12	19.12	11.76	0.00	0.00	0.00	0.00
	Total (N=111)	59.46	18.92	12.61	4.50	1.80	1.80	0.90

Table 3.4: Number of Individual Dominance Violations

violations are delivered through indirect comparisons, this is potentially suggestive of explicit costs of consideration for subjective acts. In Appendix 3.A4 we provide estimates specifically accounting for such consideration costs. In our preferred specification, Table 3.A2, column (4), consideration costs for subjective acts f and g are found to be around 4 percentage points. That is, not considering ambiguity is estimated be worth about a 4% chance of winning \$30. We reject the null hypothesis of zero consideration costs at all conventional levels, ($\chi^2(2) = 29.2,\ p < 0.01$).

3.5 Discussion and Conclusion

Subjective Expected Utility (SEU) provides a representation of choice that is often rejected by urn paradoxes of the form presented by Ellsberg (1961). Such phenomena draw attention to inconsistencies between two subjective acts. In returning to the consistency implications of SEU, our tests also draw attention to inconsistencies for subjective-objective mixtures. Analyzing subjective-objective inconsistencies is critical because non-SEU models are potentially separable along this dimension. That is, though the popularized models of Max-min Expected Utility (MEU) (Gilboa & Schmeidler, 1989) and Klibanoff et al. (2005) (KMM) both deliver inconsistencies between two subjective bets, only KMM can deliver Subjective-Objective inconsistencies. Under MEU's relaxed independence axiom, Certainty Independence, one may carry pessimistic beliefs with respect to a given act, but that pessimism must be consistently applied to all subjective-objective

low-paying marbles.

mixtures. Under KMM, mixing with any act yields hedging benefits, including mixing with objective lotteries.

We document Subjective-Subjective inconsistencies in the form of the Ellsberg (1961) two-color example by comparing the average valuation of two complementary acts, f and g, to the valuation of an objective 50-50 mixture, $\frac{1}{2}h + \frac{1}{2}l$. Subjective acts, on average, are valued around 9% below their objective counterpart and 60% of subjects are inconsistent in the Ellsberg sense. The average behavior is consistent with believing a fully ambiguous jar contains around 70% low-paying marbles regardless of whether the act in question is f or g.

The novel result of our experiment is the documentation of Subjective-Objective inconsistencies. Our results point to a "directed pessimism" wherein subjective acts are treated differently depending on whether they are mixed with high or low outcomes. Individuals do not treat ambiguous jar portions identically when mixing with high and low. Rather, the average behavior is consistent with believing the ambiguous portion is around 90% low-paying marbles when mixing with the high outcome, h, and only around 55% low-paying marbles when mixing with the low outcome, l. Further, 95% of subjects exhibit such directed pessimism and it is largely independent of standard Ellsberg-style behavior.

The collection of behavior lies outside both MEU and popularized functional forms for KMM. However, given the flexibility of the KMM model, the broad patterns could potentially be accommodated by alternative functional forms with specific restrictions on the nature of the aggregator function, $\phi(\cdot)$. Hence, our results on the collection of Subjective-Subjective and Subjective-Objective inconsistencies may helpfully distinguish both between competing non-SEU models and, within the class of smooth models, the nature of preferences.

Our individual data allow us to investigate a further pattern of inconsistency in comparisons between subjective and objective acts. In particular our data show a prominence of dominance violations, particularly for mixtures with the high outcome, h. That is, an ambiguous jar portion is treated as potentially being worse than its worst possible distribution. Such behavior lies outside of all considered SEU and non-SEU models. Though one might view such behavior as mistaken, it does correlate highly with standard Ellsberg-style behavior. Hence, we are hesitant to jump to conclusions. We present the violations as primarily suggestive, potentially indicating some costs of consideration for ambiguous acts. Importantly, such consideration costs need not be large to generate our effects. Estimates of such cost parameters are around a 4% chance of winning \$30. That is, subjects are willing to give up around \$1.20 in expected value to not have to think about ambiguity.

Given our results lie outside both SEU and current non-SEU models of choice, they present an important challenge for theoretical developments of decision-making under ambiguity. Articulating such a model is outside of the scope of this work, but our analysis points to several key elements. Mixture dependent ambiguity attitudes could be a feature of such a model, deliverable either through non-linearities as in KMM, or through explicit statement of directed pessimism. Consideration costs could be another feature, deliverable either through changing curvature under ambiguity or fixed costs. In Appendix 3.A4, we estimate a benchmark of SEU and then allow explicitly for non additive beliefs of the MEU form, mixture dependence in beliefs, and fixed consideration costs. Though far from an articulated model of choice and without even a hint of the necessary axiomatic foundation, this suggestive first step matches well the key components of the data.

Even without a model, we believe insights can be drawn for applied work relying on reduced form results. In particular, the intuition of directed pessimism may shed light on a variety of phenomena in risk taking. One natural application might be the environment of rainfall index insurance in developing countries. Take up of such products has been generally low (for discussion, see e.g., Chantarat, Mude, Barrett, & Carter, 2013). The value of index insurance can be viewed as the valuation of a subjective act and standard economics holds that low demand is driven by beliefs that make the product unattractive. Now consider offering the contract where half the time the person receives the index insurance payout and half the time they receive a high outcome regardless of the realized event. Under the view of directed pessimism, this would make individuals substantially more pessimistic for the ambiguous portion of the act and so make insurance take-up more likely. Though such a product has, to our knowledge, never been tested, it is in some ways close in spirit to a financial innovation occurring in the life insurance industry in the 19th century. Tontine life insurance was introduced at a moment of stagnation in the term life insurance industry (Ransom & Sutch, 1987). The product took the insurance premiums paid by policy holders and, instead of paying out annual dividends, invested the surplus premiums in a tontine account. At the end of the term, the tontine account was paid to survivors current on their policies. Demand increased dramatically such that by 1906, before tontine insurance became illegal, an estimated two-thirds of policies were of the form (Ransom & Sutch, 1987). In effect, the innovation made the bad state worse (as no premium surplus dividends were received) and the good state better (as one won the survivor lottery). With objective probabilities, no dependence on others, and actuarially fair values this should, in the presence of risk aversion, lead to less demand. Stylistically, the tontine policy would generate a mean preserving spread of the initial policy. Additionally, however, the tontine mixed the insurance act in question with an additional act based upon others' deaths and policy maintenance. For the healthy and organized this may be mixing with quite a good act, with directed pessimism leading to increased demand, and, interestingly, positive selection. Though these are only two examples, we believe the insights from this experiment for directed pessimism, as well as those for consideration costs, may prove helpful in other applied settings.

3.6 References

Abdellaoui, M., Baillon, A., Placido, L. L., & Wakker, P. P. (2010). The rich domain of uncertainty: source functions and their experimental implementation. *The American Economic Review*, *101*(April), 695–723.

Abdellaoui, M., Klibanoff, P., & Placido, L. (2011). *Ambiguity and compound risk attitudes :*

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole americaine. *Econometrica*, *21*(4), 503–546.

Anscombe, F. & Aumann, R. (1963). A definition of subjective probability. *The Annals of Mathematical Statistics*, *34*(1), 199–205.

Azriely, Y., Chambers, C. P., & Healy, P. J. (2014). *Incentives in experiments: a theoretical analysis.*

Bade, S. (2014). *Randomization devices and the elicitation of ambiguity averse preferences.*

Baillon, A., Halevy, Y., & Li, C. (2014). *Experimental elicitation of ambiguity attitude.*

Baillon, A. & Bleichrodt, H. (2011). *Testing ambiguity models through the measurement of probabilities for gains and losses.*

Becker, S. W. & Brownson, F. O. (1964). What price ambiguity? or the role of ambiguity in decision-making. *The Journal of Political Economy*, *72*(1), 62–73.

Binmore, K., Stewart, L., & Alex Voorhoeve. (2013). *How much ambiguity aversion? - finding indifferences between ellsberg's risky and ambiguous bets.*

Bleichrodt, H., Abellan-Perinan, J. M., Pinto-Prades, J. L., & Mendez-Martinez, I. (2007). Resolving inconsistencies in utility measurment under risk: tests of generalizations of expected utility. *Management Science*, *53*(3).

Camerer, C. & Weber, M. (1992). Recent developments in modeling preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370.

Chantarat, S., Mude, A. G., Barrett, C. B., & Carter, M. R. (2013). Designing index-based livestock insurance for managing asset risk in northern kenya. *Journal of Risk and Insurance*, *80*(1), 205–237.

Charness, G., Karni, E., & Levin, D. (2012). *Ambiguity attitudes : an experimental investigation*.

Curley, S. P. & Yates, J. F. (1985). The center and range of the probability interval as factors affecting ambiguity preferences. *Organizational Behavior and Human Decision Processes*, *36*, 273–287.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, *75*(4), 643.

Epstein, L. & Schneider, M. (2010). Ambiguity and asset markets. *NBER Working Paper*, (16181).

Farquhar, P. H. (1984). Utility assessment methods. *Management Science*, *30*(11), 1283–1300.

Ghirardato, P., Maccheroni, F., & Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, *118*(2), 133–173.

Gilboa, I. & Marinacci, M. (2011). *Ambiguity and the bayesian paradigm*.

Gilboa, I. & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*, 141–153.

Halevy, Y. (2007). Ellsberg revisited: an experimental study. *Econometrica*, *75*(2), 503–536.

Hardy, G. H., Littlewood, J. E., & Pólya, G. (1934). *Inequalities* (1991st ed.). Cambridge Mathematical Library. Cambridge, UK: Cambridge University Press.

Holt, C. (1986). Preference reversals and the independence axiom. *The American Economic Review*, *76*(1), 508–515.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, *92*(5), 1644–1655.

Hong, C. S., Bin, M., & Songfa, Z. (2013). *Partial ambiguity*.

Karni, E. & Safra, Z. (1987). "Preference reversal"and the observability of preferences by experimental methods. *Econometrica*, *55*(3), 675–685.

Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, *73*(6), 1849–1892.

Magat, W. A., Viscusi, W. K., & Huber, J. (1996). A reference lottery metric for valuing health. *Management Science*, *42*(8), 1118–1130.

McCord, M. & de Neufville, R. (1986). 'lottery equivalents': reduction of the certainty effect problem in utility assignment. *Management Science*, *32*(1), 56–60.

Meier, S. & Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal - Applied Economics*, *2*(1), 193–210.

Oliver, A. (2005). Testing the internal consistency of the lottery equivalents method using health outcomes. *Health Economics*, *14*, 149–159.

Oliver, A. (2007). A qualitative analysis of the lottery equivalents method. *Economics and Philosophy*, *23*, 185–204.

Pratt, J. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*(1-2), 122–136.

Raiffa, H. (1961). Risk, ambiguity, and the savage axioms: comment. *The Quarterly Journal of Economics*, *75*(4), 690–694.

Ransom, R. L. & Sutch, R. (1987). Tontine insurance and the armstrong investigation: a case of stifled innovation 1868–1905. *The Journal of Economic History*, *47*(2), 379–390.

Samuelson, P. A. (1953). Utilité, préfèrence et probabilité. In *Colloques internationaux du centre natlonal de la recherche scientifique (econometrie) 40* (pp. 141–150).

Savage, L. J. (1953). Une axiomatisation de comportement raisonnable face à l'incertitude. In *Colloques internationaux du centre natlonal de la recherche scientifique (econometrie) 40* (pp. 29–33).

Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley and Sons.

Schmeidler, D. (1989). Probability and expected utility without additivity. *Econometrica*, *57*(3), 571–587.

Stewart, M. B. (1983). On least squares estimation when the dependent variable is grouped. *The Review of Economic Studies*, *50*(4), 737–753.

Varian, H. R. (1992). *Microeconomic analysis* (3rd). New York: Norton.

3.7 Appendices

3.A1 KMM Preferences and $\delta_M(\alpha, f)$ / $\delta_M(\alpha, g)$

In this appendix section we demonstrate that for KMM preferences,

$$U_{KMM}(f) = \int_S p(s)\ \phi\left(u(a(s))\right)\ ds,$$

with $\phi(\cdot)$ strictly increasing and strictly concave everywhere, $\delta_M(\alpha, f) \geq 0$ for all α.

We introduce an Ellsberg style jar composed of N red and green marbles. The state of the world is determined by the composition of the jar.[48] Hence the state space has $N+1$ elements and we can label states $S = \{0, .., N\}$ by how many of the marbles in the jar are red. We establish the set of consequences as $X = \{0, x, y\}$ with $0 < x < y$. Under the state space as defined above our four basic acts are:

- f: in each state i: yields y with probability $\frac{i}{N}$, yields x with probability $\frac{N-i}{N}$.
- g: in each state i: yields y with probability $\frac{N-i}{N}$, yields x with probability $\frac{i}{N}$.
- h: in each state i: yields y with probability 1.
- l: in each state i: yields x with probability 1.

Consider the act f, which yields the KMM utility

$$U_{KMM}(f) = \sum_{i=0}^{N} \pi_i \phi\left(\frac{i}{N}u(y) + \frac{N-i}{N}u(x)\right)$$

where π_i is the subjective probability the decision maker attaches to the state where i marbles are red yielding y and $N-i$ marbles are green yielding x. Rewrite this as

$$U_{KMM}(f) = \sum_{i=0}^{N} \pi_i \phi(A_i),$$

[48] We use this state space rather than one in which the state is determined by the color of the drawn marble because while the two state space formulations are equivalent under SEU, MEU and CEU (the probabilities in the compositions-as-states formulation can be reduced to yield the draws-as-states formulation) assuming the binary state space is not without loss under KMM

where $A_i = \frac{i}{N}u(y) + \frac{N-i}{N}u(x)$. Now consider the mixture acts $\alpha f + (1-\alpha)h$ and $\alpha f + (1-\alpha)l$ with $\alpha \in (0,1)$,

$$U_{KMM}(\alpha f + (1-\alpha)h) = \sum_{i=0}^{N} \pi_i \phi(\alpha A_i + (1-\alpha)u(y)),$$

$$U_{KMM}(\alpha f + (1-\alpha)l) = \sum_{i=0}^{N} \pi_i \phi(\alpha A_i + (1-\alpha)u(x)).$$

Note, by the strict concavity of $\phi(\cdot)$

$$\phi(\alpha A_i + (1-\alpha)u(y)) > \alpha\phi(A_i) + (1-\alpha)\phi(u(y))$$

$$\phi(\alpha A_i + (1-\alpha)u(x)) > \alpha\phi(A_i) + (1-\alpha)\phi(u(x))$$

Let $\epsilon_{yA_i} > 0$ and $\epsilon_{xA_i} > 0$ be defined such that

$$\phi(\alpha A_i + (1-\alpha)u(y)) = \alpha\phi(A_i) + (1-\alpha)\phi(u(y)) + \epsilon_{yA_i}$$

$$\phi(\alpha A_i + (1-\alpha)u(x)) = \alpha\phi(A_i) + (1-\alpha)\phi(u(x)) + \epsilon_{xA_i}$$

Hence $U_{KMM}(\alpha f + (1-\alpha)h) - U_{KMM}(\alpha f + (1-\alpha)l)$ is

$$\sum_{i=0}^{N} \pi_i \{[\alpha\phi(A_i) + (1-\alpha)\phi(u(y)) + \epsilon_{yA_i}] - [\alpha\phi(A_i) + (1-\alpha)\phi(u(x)) + \epsilon_{xA_i}]\},$$

or

$$(1-\alpha)\phi(u(y)) - (1-\alpha)\phi(u(x)) + \sum_{i=0}^{N} \pi_i[\epsilon_{yA_i} - \epsilon_{xA_i}].$$

Now consider the mixture act $\alpha h + (1-\alpha)l$ with corresponding KMM utility

$$U_{KMM}(\alpha h + (1-\alpha)l) = \phi(\alpha u(y) + (1-\alpha)u(x)),$$

Note, by the strict concavity of $\phi(\cdot)$

$$\phi(\alpha y + (1-\alpha)u(x)) > \alpha\phi(u(y)) + (1-\alpha)\phi(u(x))$$

Let ϵ be defined such that

$$\phi(\alpha y + (1-\alpha)u(x)) = \alpha\phi(u(y)) + (1-\alpha)\phi(u(x)) + \epsilon.$$

Hence,

$$U_{KMM}(\alpha h + (1-\alpha)l) = \alpha\phi(u(y)) + (1-\alpha)\phi(u(x)) + \epsilon.$$

Now consider $U_{KMM}(\alpha f+(1-\alpha)h) - U_{KMM}(\alpha f+(1-\alpha)l) + U_{KMM}(\alpha h+(1-\alpha)l)$. Under SEU such a sum is equal to the utility of the highest outcome, y, or, by normalization equal to 1. Under KMM this sum is

$$(1-\alpha)\phi(u(y)) - (1-\alpha)\phi(u(x)) + \alpha\phi(y) + (1-\alpha)\phi(u(x)) + \epsilon + \sum_{i=0}^{N} \pi_i[\epsilon_{yA_i} - \epsilon_{xA_i}],$$

or

$$\phi(u(y)) + \sum_{i=0}^{N} \pi_i[\epsilon_{yA_i} - \epsilon_{xA_i} + \epsilon],$$

Normalizing $\phi(u(y)) = 1$ this yields

$$1 + \sum_{i=0}^{N} \pi_i[\epsilon_{yA_i} - \epsilon_{xA_i} + \epsilon],$$

We now recall the definition of $\delta_M(\alpha, f)$ as

$$\delta_M(\alpha, f) = U(\alpha f + (1-\alpha)h) - U(\alpha f + (1-\alpha)l) + U(\alpha h + (1-\alpha)l) - 1.$$

Subtracting 1 from the above we obtain only the sum of error terms,

$$\sum_{i=0}^{N} \pi_i[\epsilon_{yA_i} - \epsilon_{xA_i} + \epsilon].$$

We can now investigate the sign of $\epsilon_{yA_i} - \epsilon_{xA_i} + \epsilon$ to recover whether $\delta_M(\alpha, f) \geq 0$ or $\delta_M(\alpha, f) \leq 0$. We wish to prove that if $\phi(\cdot)$ is concave, $\epsilon_{yA_i} - \epsilon_{xA_i} + \epsilon \geq 0$ for every A_i and for every α such that $\delta_M(\alpha, f) \geq 0$ for every α.

Note that for a given $u(x) < A_i < u(y)$, the error terms are each associated with the distance from a chord between two points and the functions $\phi(\cdot)$. The location

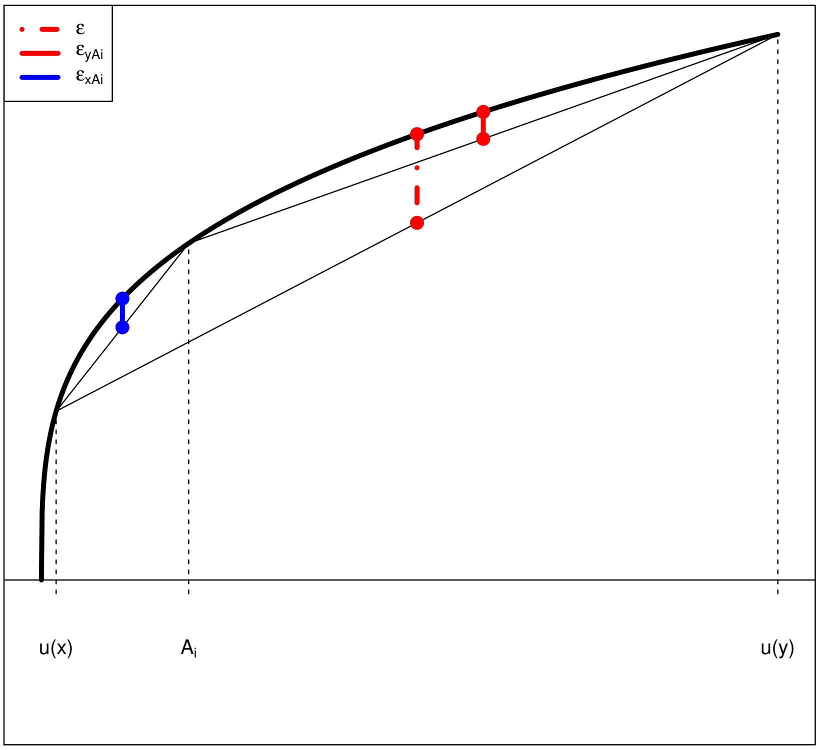


Figure 3.A1: A Standard Case

on the chord is determined by the mixture α. Figure 3.A1 illustrates a standard case with ϵ_{yA_i} and ϵ in red and ϵ_{xA_i} in blue. We can investigate the limiting case that maximizes ϵ_{xA_i} and minimizes ϵ_{yA_i} and ϵ. This is done by replacing the function away from $u(x)$ with the extended chord between A_i and $u(y)$. This is illustrated in Figure 3.A2.

In the limiting case, $\epsilon_{yA_i} = 0$. Hence, the relevant comparison is between ϵ_{xA_i} and ϵ. Note that the slope of the extended chord in Figure 3.A2 is

$$\frac{\phi(u(y)) - \phi(A_i)}{u(y) - A_i}.$$

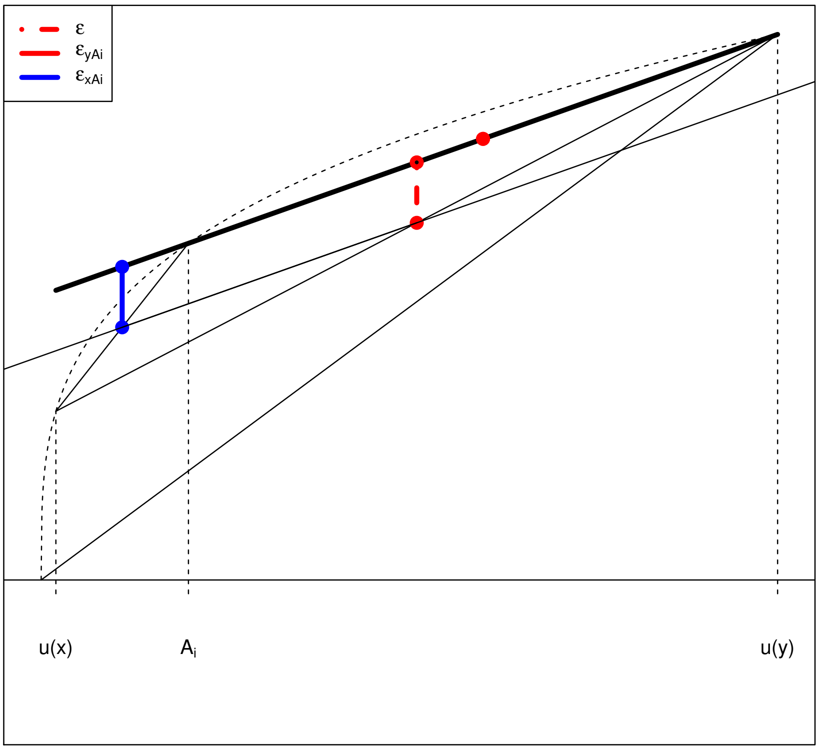


Figure 3.A2: Limiting Case

Now consider the slope between the points, $(\alpha A_i + (1-\alpha)u(x),\ \alpha\phi(A_i) + (1-\alpha)\phi(u(x)))$ and $((\alpha u(y) + (1-\alpha)u(x),\ \alpha\phi(u(y) + (1-\alpha)\phi(u(x)))$. This slope is

$$\frac{[\alpha\phi(u(y) + (1-\alpha)\phi(u(x))] - [\alpha\phi(A_i) + (1-\alpha)\phi(u(x))]}{[\alpha u(y) + (1-\alpha)u(x)] - [\alpha A_i + (1-\alpha)u(x)]} = \frac{\phi(u(y)) - \phi(A_i)}{u(y) - A_i},$$

parallel to the extended chord. Hence, a parallelogram is constructed with the error terms, ϵ and ϵ_{xA_i} and the chords connecting their endpoints. Because in a parallelogram opposite sides are of equal length, in the limiting case $\epsilon = \epsilon_{xAi}$.

As $\phi(\cdot)$ is assumed strictly concave everywhere, the limiting case occurs only if $i = 0$ or $i = N$. In all other cases ϵ is strictly bounded below by its limiting value and ϵ_{xAi} is strictly bounded above by its limiting value. Hence, $\epsilon \geq \epsilon_{xAi}$ always and $\epsilon_{yAi} - \epsilon_{xAi} + \epsilon \geq 0$. The same graphical technique can be applied for every A_i and for every α. Hence, for a strictly concave and increasing $\phi(\cdot)$ we conclude that $\epsilon_{yA_i} - \epsilon_{xA_i} + \epsilon \geq 0$ for every A_i and for every α. Moreover, for all $0 < i < N$ the inequality is strict. For strictly concave and strictly increasing $\phi(\cdot)$, $\delta_M(\alpha, f) \geq 0$ for every α. One can easily prove the opposite for a strictly convex and strictly increasing $\phi(\cdot)$.

3.A2 KMM Preferences and $d_M(\alpha, f)$ / $d_M(\alpha, g)$

In this appendix section we prove results about the shape of the uncertainty equivalent functions and the empirical mixture differences $d_M(\alpha, f)$ and $d_M(\alpha, g)$ under KMM. The setting is identical to the setting assumed in the previous section.

Theorem 1. *For a KMM agent with strictly increasing and four times continuously differentiable utility aggregator $\phi(\cdot)$ uncertainty equivalents for the subjective-objective mixtures $\alpha f + (1-\alpha)(pl + (1-p)h)$ with $\alpha \in (0,1)$ and $p \in [0,1]$ are*

- *a concave function of the mixture probability α if $\phi''(\cdot) < 0$ and the coefficient of absolute ambiguity tolerance $\frac{-\phi'(\cdot)}{\phi''(\cdot)}$ is convex, and*
- *a convex function of the mixture probability α if $\phi''(\cdot) > 0$ and the coefficient of absolute ambiguity tolerance $\frac{-\phi'(\cdot)}{\phi''(\cdot)}$ is concave.*

Proof. The proof is adapted from the proof of Theorem 106 in Hardy, Littlewood, and Pólya (1934). The uncertainty equivalent q for the subjective-objective mixture $\alpha f + (1-\alpha)(pl + (1-p)h)$ satisfies the following equality:

$$\phi\left(qu(y) + (1-q)\,u(0)\right) = \sum_{i=0}^{N} \pi_i \phi \Bigg(\alpha \left(\frac{i}{N} u(y) + \left(1 - \frac{i}{N}\right) u(x) \right) \\ + (1-\alpha)\left((1-p)\,u(y) + pu(x)\right) \Bigg)$$

Since $\phi(\cdot)$ is strictly increasing, it is invertible. Therefore we can write q explicitly as

$$q = \frac{\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(\alpha\left(\frac{i}{N}u(y) + \left(1-\frac{i}{N}\right)u(x)\right) + (1-\alpha)\left((1-p)\,u(y) + pu(x)\right)\right)\right) - u(0)}{u(y) - u(0)}$$

In the following define $p_i \equiv \frac{i}{N} + p - 1$ and $E_i \equiv u(x) + (\alpha p_i + (1-p))\,(u(y) - u(x))$.

The last expression can then be written more concisely as

$$q = \frac{\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right) - u(0)}{u(y) - u\left(0\right)}$$

Taking partial derivatives w.r.t α twice:

$$\begin{aligned}
\frac{\partial q}{\partial \alpha} =& \frac{u(y)-u(x)}{u(y)-u\left(0\right)} \times \left(\phi^{-1}\right)^{'}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\left[\sum_{i=0}^{N}\pi_i\phi^{'}\left(E_i\right)p_i\right] \\
\frac{\partial^2 q}{\partial \alpha^2} =& \frac{\left(u(y)-u(x)\right)^2}{u(y)-u\left(0\right)} \\
& \times \Bigg[\left(\phi^{-1}\right)^{''}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\left[\sum_{i=0}^{N}\pi_i\phi^{'}\left(E_i\right)p_i\right]^2 \\
& + \left(\phi^{-1}\right)^{'}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\left[\sum_{i=0}^{N}\pi_i\phi^{''}\left(E_i\right)p_i^2\right]\Bigg] \\
=& \frac{\left(u(y)-u(x)\right)^2}{u(y)-u\left(0\right)} \times \Bigg[\frac{-\phi^{''}\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)}{\left[\phi^{'}\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)\right]^3}\left[\sum_{i=0}^{N}\pi_i\phi^{'}\left(E_i\right)p_i\right]^2 \\
& + \frac{1}{\phi^{'}\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)}\left[\sum_{i=0}^{N}\pi_i\phi^{''}\left(E_i\right)p_i^2\right]\Bigg]
\end{aligned}$$

A necessary and sufficient condition for global concavity of q is that the second derivative is non-positive throughout, i.e. that

$$\begin{aligned}
&\left[\phi^{'}\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)\right]^2\left[\sum_{i=0}^{N}\pi_i\phi^{''}\left(E_i\right)p_i^2\right] \\
&- \phi^{''}\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)\left[\sum_{i=0}^{N}\pi_i\phi^{'}\left(E_i\right)p_i\right]^2 \leq 0
\end{aligned}$$

or equivalently that

$$\frac{\left[\phi'\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)\right]^2}{\phi''\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)} \leq \frac{\left[\sum_{i=0}^{N}\pi_i\phi'\left(E_i\right)p_i\right]^2}{\sum_{i=0}^{N}\pi_i\phi''\left(E_i\right)p_i^2}$$

The numerator of the right-hand side expression can be bounded above by the Cauchy-Schwartz inequality:

$$\begin{aligned}\left[\sum_{i=0}^{N}\pi_i\phi'\left(E_i\right)p_i\right]^2 &= \left[\sum_{i=0}^{N}\sqrt{\phi''\left(E_i\right)\pi_i}p_i\sqrt{\frac{\pi_i\left(\phi'\left(E_i\right)\right)^2}{\phi''\left(E_i\right)}}\right]^2 \\ &\leq \left[\sum_{i=0}^{N}\phi''\left(E_i\right)\pi_i p_i^2\right]\cdot\left[\sum_{i=0}^{N}\pi_i\frac{\left(\phi'\left(E_i\right)\right)^2}{\phi''\left(E_i\right)}\right]\end{aligned}$$

Dividing this inequality by the denominator of the original expression yields a lower bound:

$$\begin{aligned}\frac{\left[\phi'\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)\right]^2}{\phi''\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)} &\leq \frac{\left[\sum_{i=0}^{N}\phi''\left(E_i\right)\pi_i p_i^2\right]\cdot\left[\sum_{i=0}^{N}\pi_i\frac{\left(\phi'\left(E_i\right)\right)^2}{\phi''\left(E_i\right)}\right]}{\sum_{i=0}^{N}\pi_i\phi''\left(E_i\right)p_i^2} \\ &\leq \frac{\left[\sum_{i=0}^{N}\pi_i\phi'\left(E_i\right)p_i\right]^2}{\sum_{i=0}^{N}\pi_i\phi''\left(E_i\right)p_i^2}\end{aligned}$$

The first inequality is therefore a sufficient condition for the second. After cancelling common terms we are left with

$$\frac{\left[\phi'\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)\right]^2}{\phi''\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(E_i\right)\right)\right)} \leq \sum_{i=0}^{N}\pi_i\frac{\left(\phi'\left(E_i\right)\right)^2}{\phi''\left(E_i\right)}$$

Define $K_i = \phi(E_i)$ and

$$\Phi(K_i) = \frac{(\phi'(\phi^{-1}(K_i)))^2}{\phi''(\phi^{-1}(K_i))}$$

and rewrite the above as

$$\Phi(\sum \pi_i K_i) \leq \sum \pi_i \Phi(K_i)$$

The *convexity* of $\Phi(\cdot)$ is therefore sufficient for the concavity of the uncertainty equivalents.

The first two derivatives of $\Phi(\cdot)$ are

$$\begin{aligned}
\Phi'(K) =& \frac{2\phi'(E)\phi''(E)\frac{dE}{dK}\phi''(E) - (\phi'(E))^2\phi'''(E)\frac{dE}{dK}}{(\phi''(E))^2} \\
=& \frac{d}{dE}\left(\frac{\phi'(E)}{\phi''(E)}\right)\phi'(E)\frac{dE}{dK} + \frac{\phi'(E)(\phi''(E))^2}{\phi''(E))^2}\frac{dE}{dK} \\
=& \frac{d}{dE}\left(\frac{\phi'(E)}{\phi''(E)}\right) + 1
\end{aligned}$$

$$\Phi''(K) = \frac{1}{\phi'(E)}\frac{d^2}{dE^2}\left(\frac{\phi'(E)}{\phi''(E)}\right)$$

$\Phi(\cdot)$ is convex if its second derivative is positive, i.e. if $\frac{\phi'(E)}{\phi''(E)}$ is convex, which is equivalent to the coefficient of absolute risk tolerance $\frac{-\phi'(E)}{\phi''(E)}$ being concave. The proof for the case when $\phi''(\cdot) > 0$ and the coefficient of absolute ambiguity tolerance $\frac{-\phi'(\cdot)}{\phi''(\cdot)}$ is convex is analogous and therefore omitted. $\square$

How restrictive is the condition that the coefficient of absolute ambiguity tolerance $\frac{-\phi'(\cdot)}{\phi''(\cdot)}$ be concave? Not very restrictive at all. In fact the whole class of HAAA (hyperbolic absolute ambiguity aversion) functions of the form

$$\phi(x) = \frac{1-\gamma}{\gamma}\left(\frac{\alpha x}{1-\gamma} + \beta\right)^\gamma$$

has coefficients of absolute ambiguity tolerance that are linear and therefore both convex and concave. The function above nests the popular exponential $\phi(x) = -e^{-\theta x}$ and power $\phi(x) = \frac{x^{1-\theta}}{1-\theta}$ functions.

We next prove a result about the sign of the empirical mixture differences $\delta_M(\alpha, f)$ and $\delta_M(\alpha, g)$, which is our main object of interest. This sign will depend upon the difference between the sign of the difference between the uncertainty equivalents of the two mixtures, which will in turn depend upon whether uncertainty equivalents as a function of α are more or less concave when mixing with h or mixing with l (for an illustration, see Figure 3.A3). Intuitively mixing with any constant act has three effects in KMM:

1. It pulls the distribution of state expected utilities in the direction of the expected utility of the constant act. This is the only effect that plays a role in SEU.
2. It reduces the spread of state expected utilities. As shown in the previous theorem, under fairly general assumptions ambiguity averse KMM agents, in contrast to SEU agents, value this reduction because state expected utilities are aggregated with the concave function $\phi(\cdot)$. This increase in utility is reflected in an additional increase in uncertainty equivalents.
3. There may be an additional effect on the uncertainty equivalents that depends on the exact functional form of $\phi(\cdot)$ in a way that is analogous to the behavior of risk premia in expected utility under risk. Just as risk premia depend on the Arrow-Pratt coefficient of absolute risk aversion (See Pratt, 1964, , Theorem 2) uncertainty equivalents depend on the coefficient of ab-

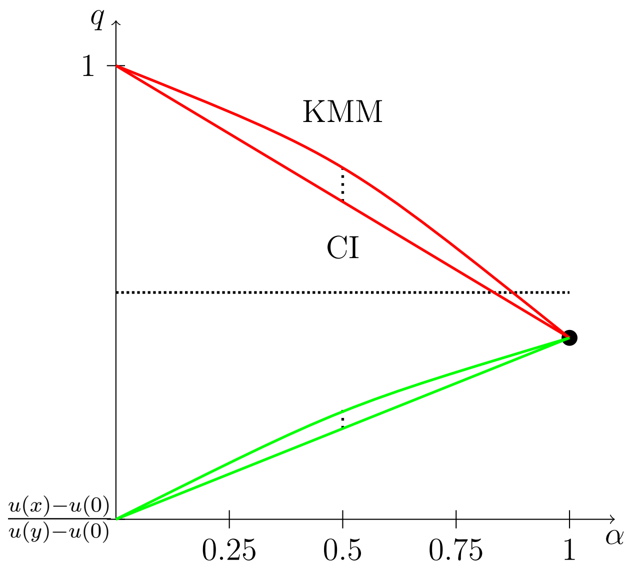


Figure 3.A3: Uncertainty equivalents q under Certainty Independence (CI) and KMM with decreasing absolute ambiguity aversion. Mixtures displayed are those with the acts h and l

solute ambiguity aversion. If the coefficient of absolute ambiguity aversion $\frac{\phi''(u)}{\phi'(u)}$ is decreasing in u, for example, this means mixing with h pulls the distribution of state expected utilities into a region in which $\phi(\cdot)$ is less concave, which raises the uncertainty equivalent further. Conversely, in this case mixing with l pulls the distribution of state expected utilities into a region in which $\phi(\cdot)$ is *more* concave, which lowers the uncertainty equivalent.

Theorem 2. *Assume a KMM agent with strictly increasing and twice continuously differentiable utility aggregator $\phi(\cdot)$.*

- *$d_m(\alpha, f) > 0$ and $d_m(\alpha, g) > 0$ for all α if the coefficient of absolutel ambiguity aversion $\frac{-\phi''(u)}{\phi'(u)}$ is strictly decreasing in u*
- *$d_m(\alpha, f) = 0$ and $d_m(\alpha, g) = 0$ for all α if the coefficient of absolutel ambiguity aversion $\frac{-\phi''(u)}{\phi'(u)}$ is constant*
- *$d_m(\alpha, f) < 0$ and $d_m(\alpha, g) < 0$ for all α if the coefficient of absolutel ambiguity aversion $\frac{-\phi''(u)}{\phi'(u)}$ is strictly increasing in u*

Proof. By definition the mixture difference $d_m(\alpha, f)$ is given by the following linear combination of uncertainty equivalents:

$$d_m(\alpha, f) = q_{\alpha f+(1-\alpha)h} - q_{\alpha f+(1-\alpha)l} + q_{\alpha h+(1-\alpha)l} - 1$$

Under KMM uncertainty equivalents of the mixtures $\alpha f + (1-\alpha)c$ where $c \in \{l, h\}$ denotes the two objective/constant acts can be written as

$$q_{\alpha f+(1-\alpha)a} = \frac{\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(\alpha U(f(i)) + (1-\alpha)U(c)\right)\right) - u(0)}{u(y) - u(0)}$$

where $U(c) = u(x)$ for act l and $U(c) = u(y)$ for act h are the von-Neumann-Morgenstern utilities of the objective acts and $U(f(i)) = \frac{i}{N}u(y) + \left(1 - \frac{i}{N}\right)u(x)$ is the vNM-utility of the lottery that f yields in state i. The uncertainty equivalent of $\alpha h + (1-\alpha)l$ is simply

$$q_{\alpha h+(1-\alpha)l} = \frac{\alpha u(y) + (1-\alpha)u(x) - u(0)}{u(y) - u(0)}$$

Substituting in these uncertainty equivalents yields

$$
\begin{aligned}
d_m(\alpha, f) =& \frac{\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(\alpha U(f(i)) + (1-\alpha)u(y)\right)\right) - u(0)}{u(y) - u(0)} \\
&- \frac{\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(\alpha U(f(i)) + (1-\alpha)u(x)\right)\right) - u(0)}{u(y) - u(0)} \\
&+ \frac{\alpha u(y) + (1-\alpha)u(x) - u(0)}{u(y) - u(0)} - 1
\end{aligned}
$$

or, after some rearrangement

$$
\begin{aligned}
d_m(\alpha, f) =& \frac{1}{u(y) - u(0)} \times \\
& \left[\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(\alpha U(f(i)) + (1-\alpha)u(y)\right)\right)\right. \\
& - \alpha\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(U(f(i))\right)\right) - (1-\alpha)u(y) \\
& - \phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(\alpha U(f(i)) + (1-\alpha)u(x)\right)\right) \\
& \left. + \alpha\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(U(f(i))\right)\right) + (1-\alpha)u(x)\right]
\end{aligned}
$$

Define $\Delta(U(c))$ as

$$
\begin{aligned}
\Delta(U(c)) =& \phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(\alpha U(f(i)) + (1-\alpha)U(c)\right)\right) \\
& - \alpha\phi^{-1}\left(\sum_{i=0}^{N} \pi_i \phi\left(U(f(i))\right)\right) - (1-\alpha)U(c)
\end{aligned}
$$

and take the partial derivative with respect to $U(c)$:

$$\begin{aligned}\frac{\partial\Delta(U(c))}{\partial U(c)} &= (1-\alpha)\Bigg[\left(\phi^{-1}\right)'\left(\sum_{i=0}^{N}\pi_i\phi\left(\alpha U(f(i))+(1-\alpha)U(c)\right)\right)\\ &\qquad\qquad\left(\sum_{i=0}^{N}\pi_i\phi'\left(\alpha U(f(i))+(1-\alpha)U(c)\right)\right)-1\Bigg]\\ &= (1-\alpha)\left[\frac{\sum_{i=0}^{N}\pi_i\phi'\left(\alpha U(f(i))+(1-\alpha)U(c)\right)}{\phi'\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(\alpha U(f(i))+(1-\alpha)U(c)\right)\right)\right)}-1\right]\end{aligned}$$

$\frac{\partial\Delta(U(c))}{\partial U(c)} > 0$ if

$$\sum_{i=0}^{N}\pi_i\phi'\left(\alpha U(f(i))+(1-\alpha)U(c)\right) > \phi'\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(\alpha U(f(i))+(1-\alpha)U(c)\right)\right)\right)$$

that is, if the average marginal utility of a mixture is larger than the marginal utility at the certainty equivalent of that mixture. To see that this is true if $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$ is decreasing, consider the indifference condition for the certainty equivalent:

$$\sum_{i=0}^{N}\pi_i\phi\left(\alpha U(f(i))+(1-\alpha)U(c)\right) = \phi\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(\alpha U(f(i))+(1-\alpha)U(c)\right)\right)\right)$$

Now modify both the mixture and its certainty equivalent by adding an $\varepsilon > 0$ to the vNM utility (not to the prize) in every state. Because the function has decreasing absolute ambiguity aversion adding this additional utility means she will now prefer the modified mixture over the modified certainty equivalent[49]:

$$\sum_{i=0}^{N}\pi_i\phi\left(\alpha U(f(i))+(1-\alpha)U(c)+\varepsilon\right) > \phi\left(\phi^{-1}\left(\sum_{i=0}^{N}\pi_i\phi\left(\alpha U(f(i))+(1-\alpha)U(c)\right)\right)+\varepsilon\right)$$

[49] The proof of this claim is analogous to the proof that risk premia in choice under risk are decreasing functions of wealth if the coefficient of absolute risk aversion is decreasing (see Pratt, 1964), which implies that such a DARA agent will prefer a risky gamble to its certainty equivalent after all prizes in the gamble and the certainty equivalent are increased by a small amount

But this must mean that

$$\sum_{i=0}^{N} \pi_i \phi' \left(\alpha U(f(i)) + (1-\alpha) U(c)\right) > \phi' \left(\phi^{-1} \left(\sum_{i=0}^{N} \pi_i \phi \left(\alpha U(f(i)) + (1-\alpha) U(c)\right) \right) \right)$$

as desired.

The cases for $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$ constant and $\frac{-\phi''(\cdot)}{\phi'(\cdot)}$ increasing are analoguous. $\square$

Note that the above result is independent of the sign of $\phi''(\cdot)$, i.e. of whether the agent is ambiguity averse or ambiguity seeking.

3.A3 Full Regression Results

Task	Act	Midpoints		Interval Regression							
		mean	s.e.	All		Ambiguity Seeking $d_{S,i} > 0$		Ambiguity Neutral $d_{S,i} = 0$		Ambiguity Averse $d_{S,i} < 0$	
1	$\frac{1}{4}g + \frac{3}{4}h$	83.90	(1.17)	83.90	(1.17)	87.00	(1.54)	88.37	(2.22)	81.47	(1.62)
2	$\frac{1}{4}g + \frac{3}{4}l$	66.91	(1.62)	66.91	(1.61)	68.00	(2.93)	73.80	(3.87)	64.26	(2.02)
3	$\frac{1}{2}g + \frac{1}{2}h$	76.46	(1.21)	76.46	(1.21)	78.00	(1.97)	83.59	(2.62)	73.60	(1.55)
4	$\frac{1}{2}g + \frac{1}{2}l$	68.40	(1.62)	68.40	(1.61)	71.50	(2.85)	75.54	(3.58)	65.07	(2.07)
5	$\frac{1}{2}g + \frac{1}{2}(\frac{1}{2}h + \frac{1}{2}l)$	73.72	(1.40)	73.72	(1.40)	76.50	(2.37)	79.46	(3.08)	70.96	(1.82)
6	$\frac{3}{4}g + \frac{1}{4}l$	68.85	(1.76)	68.85	(1.76)	73.50	(2.79)	77.50	(3.35)	64.56	(2.35)
7	$\frac{3}{4}g + \frac{1}{4}h$	74.12	(1.44)	74.12	(1.44)	78.00	(2.30)	82.93	(2.31)	70.00	(1.93)
8	g	69.80	(1.61)	69.80	(1.60)	76.00	(2.46)	78.80	(3.06)	64.93	(2.08)
9	$\frac{1}{4}l + \frac{3}{4}h$	84.71	(0.84)	84.71	(0.84)	79.00	(2.30)	87.28	(1.70)	85.51	(0.95)
10	$\frac{1}{2}l + \frac{1}{2}h$	76.37	(1.20)	76.37	(1.20)	69.25	(2.39)	78.80	(2.90)	77.65	(1.45)
11	$\frac{3}{4}l + \frac{1}{4}h$	68.76	(1.46)	68.76	(1.45)	66.25	(2.61)	73.80	(3.73)	67.79	(1.80)
12	l	62.41	(1.77)	62.41	(1.76)	60.00	(3.34)	70.76	(4.49)	60.29	(2.13)
13	$\frac{1}{4}f + \frac{3}{4}h$	84.39	(0.78)	84.39	(0.78)	84.50	(1.52)	89.02	(1.81)	82.79	(0.94)
14	$\frac{1}{4}f + \frac{3}{4}l$	66.15	(1.66)	66.15	(1.66)	69.00	(2.82)	72.28	(4.02)	63.24	(2.10)
15	$\frac{1}{2}f + \frac{1}{2}h$	77.09	(1.19)	77.09	(1.19)	80.25	(2.02)	83.59	(2.33)	73.97	(1.54)
16	$\frac{1}{2}f + \frac{1}{2}l$	67.36	(1.73)	67.36	(1.72)	72.00	(2.79)	75.33	(3.65)	63.31	(2.24)
17	$\frac{1}{2}f + \frac{1}{2}(\frac{1}{2}h + \frac{1}{2}l)$	73.09	(1.35)	73.09	(1.34)	76.00	(2.44)	79.46	(2.94)	70.07	(1.71)
18	$\frac{3}{4}f + \frac{1}{4}l$	68.63	(1.69)	68.63	(1.68)	72.25	(3.13)	77.72	(3.28)	64.49	(2.17)
19	$\frac{3}{4}f + \frac{1}{4}h$	73.09	(1.50)	73.09	(1.49)	77.25	(3.40)	81.20	(2.73)	69.12	(1.86)
20	f	69.89	(1.71)	69.89	(1.70)	77.50	(2.16)	78.80	(2.97)	64.63	(2.28)
d_S		-6.53	(1.14)	-6.53	(1.13)	7.50	(1.48)			-12.87	(1.24)
$H_0 : d_S = 0$		$F = 33.10$ ($p < 0.01$)		$\chi^2(1) = 33.39$ ($p < 0.01$)		$\chi^2(1) = 25.85$ ($p < 0.01$)				$\chi^2(1) = 107.00$ ($p < 0.01$)	
$d_M(g, \frac{1}{4})$		-14.26	(1.37)	-14.26	(1.36)	-14.75	(2.34)	-11.63	(2.22)	-15.00	(1.97)
$d_M(g, \frac{1}{2})$		-15.56	(1.16)	-15.56	(1.16)	-24.25	(2.28)	-13.15	(2.19)	-13.82	(1.47)
$d_M(g, \frac{3}{4})$		-10.02	(1.06)	-10.02	(1.06)	-16.50	(2.70)	-7.283	(1.87)	-9.044	(1.31)
$d_M(f, \frac{1}{4})$		-13.00	(1.11)	-13.00	(1.10)	-18.25	(2.69)	-9.457	(2.65)	-12.65	(1.27)
$d_M(f, \frac{1}{2})$		-13.90	(1.27)	-13.90	(1.26)	-22.50	(2.97)	-12.93	(2.25)	-11.69	(1.59)
$d_M(f, \frac{1}{4})$		-10.83	(1.17)	-10.83	(1.17)	-16.00	(3.28)	-9.239	(1.74)	-9.853	(1.49)
$H_0 : d_M(f, \alpha) = d_M(g, \alpha) = 0;$ $\alpha \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$				$\chi^2(6) = 261.50$ ($p < 0.01$)		$\chi^2(6) = 184.94$ ($p < 0.01$)		$\chi^2(6) = 45.82$ ($p < 0.01$)		$\chi^2(6) = 155.89$ ($p < 0.01$)	
$H_0 : d_M(f, \alpha), d_M(g, \alpha)$ identical for $d_{S,i} > 0$ and $d_{S,i} < 0$						$\chi^2(6) = 22.30$ ($p < 0.01$)					
$H_0 : d_M(f, \alpha), d_M(g, \alpha)$ identical for $d_{S,i} = 0$ and $d_{S,i} < 0$								$\chi^2(6) = 4.34$ ($p = 0.63$)			
# Individuals		111		111		20		23		68	

Interval regression. Standard errors clustered on individual level. Subjective premium and mixture distances calculated as linear combinations of regression coefficients. Hypothesis tests across $d_{S,i}$ groups from regression of response on indicators for task number interacted with $d_{S,i}$ group.

Table 3.A1: Mean uncertainty equivalents, subjective premia and mixture distances.

3.A4 Estimating Decisionmaking Parameters

Our data are clearly at odds with SEU as well as popular non-SEU formulations such as MEU and KMM. Our findings of Subjective-Subjective and Subjective-Objective inconsistency suggest two key features of decision-making: non-unique beliefs across subjective acts and a sensitivity of either the perception of ambiguity or subjects' attitude towards it to mixing subjective acts with either high or low outcomes. In this appendix we attempt to explicitly model these two features and provide corresponding parameter estimates. We additionally provide estimates related to consideration costs for ambiguous acts that may account for the observed violations of dominance. In the spirit of full disclosure, providing simple parametric methods for the estimation of beliefs and risk aversion in SEU and non-SEU models was an original objective in experimental design. Given the non-adherence of the data to all considered models of decision making under ambiguity, this exercise moved from being at the fore to an exploration outside of the theory.

For a given individual who makes $j = 1, 2, \ldots, J$ decisions, let $\tilde{p}_{green}(f)$ be the beliefs on the probability of a green marble induced by subjective act f, where a green marble yields x and a red marble yields $y > x$. Consider binary objective acts, c_j, with the same consequences as f, summarized by p_{cj} and mixtures $\alpha_j \in [0, 1]$ representing the degree of ambiguity. Let preferences satisfy SEU such that the utility of the mixture $\alpha_j f + (1 - \alpha_j)c_j$ is represented as

$$\alpha_j \cdot [\tilde{p}_{green}(f)u(x) + (1 - \tilde{p}_{green}(f))u(y)] + (1 - \alpha_j) \cdot [p_{cj}u(x) + (1 - p_{cj})u(y)].$$

We consider the equivalent constant act q_j over y and 0, summarized by q_j such that the indifference condition

$$\alpha_j \cdot [\tilde{p}_{green}(f)u(x) + (1 - \tilde{p}_{green}(f))u(y)] + (1 - \alpha_j) \cdot [p_{cj}u(x) + (1 - p_{cj})u(y)] = q_j u(y) + (1 - q_j)u(0)$$

is met. Rearranging terms one arrives at the equality

$$q_j = 1 + \alpha_j \cdot \tilde{p}_{green}(f) \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]} + (1 - \alpha_j) \cdot p_{cj} \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]}.$$

This in turn suggests a method for simultaneously estimating and testing the consistency of beliefs, $\tilde{p}_{green}(f)$, and risk aversion, $[u(x)-u(y)]/[u(y)-u(0)]$. Variation in mixtures α_j alone can identify a combination of beliefs and risk aversion, while for a fixed α_j, variation in p_{cj} identifies risk aversion independently.[50]

We assume an individual makes her J choices following

$$q_j = 1+\alpha_j \cdot \tilde{p}_{green}(f) \cdot \frac{[u(x)-u(y)]}{[u(y)-u(0)]} + (1-\alpha_j) \cdot p_{cj} \cdot \frac{[u(x)-u(y)]}{[u(y)-u(0)]} + u_j, \quad j = 1, \ldots, J, \tag{3.3}$$

where u_j is assumed to be drawn identically and independently from a mean zero normal distribution with variance σ_j^2. Define the vector $\mathbf{x}_j = [1,\ \alpha_j,\ (1-\alpha_j)p_{cj}]$ such that one can rewrite (3) as

$$q_j = \mathbf{x}_j \beta + u_j \tag{3.4}$$

where $\beta = [1,\ \beta_1,\ \beta_2]' = [1,\ \tilde{p}_{green}(f) \cdot \frac{[u(x)-u(y)]}{[u(y)-u(0)]},\ \frac{[u(x)-u(y)]}{[u(y)-u(0)]}]'$

Importantly, in our experimental environment, we do not observe q_j directly as individual responses are interval coded. Maximum likelihood methods such as those outlined in Stewart (1983) can be implemented to consistently estimate the key parameters of interest. We observe responses in one of twenty intervals $(A_{k-1}, A_k) = (0, 0.05), (0, 05, 0.10), \ldots, (0.95, 1)$. We assess the probability of an observation lying within a given interval as

$$\Phi(\frac{A_k - \mathbf{x}_j\beta}{\sigma_j}) - \Phi(\frac{A_{k-1} - \mathbf{x}_j\beta}{\sigma_j}),$$

where $\Phi(\cdot)$ represents the standard normal distribution. Hence the log-likelihood

[50] Note that in (3) variation in mixtures α_j identify a combination of beliefs, $\tilde{p}_{green}(f)$, and risk aversion for ambiguous outcomes, $[u(x)-u(y)]/[u(y)-u(0)]$. For a fixed α_j, variation in p_{cj} identifies risk aversion independently, though for unambiguous outcomes. This implies that without the assumption that risk aversion is the same across the two, one cannot separately identify beliefs and risk aversion for ambiguous outcomes. Further, without this assumption one cannot establish that the constant is equal to 1 in (3). However, even without this assumption, one can potentially still make progress exploring differential behavior across conditions. One can restate the comparative statics on beliefs presented in Table 3.A2 as comparative statics on risk aversion under ambiguity, fixing some level of belief. We opt to maintain the assumption that $[u(x) - u(y)]/[u(y) - u(0)]$ is constant over ambiguous and unambiguous outcomes and present our analysis in terms of comparative statics on beliefs.

for a given individual is

$$\sum_{j=1}^{J} log[\Phi(\frac{A_k - \mathbf{x}_j\beta}{\sigma}) - \Phi(\frac{A_{k-1} - \mathbf{x}_j\beta}{\sigma})]. \tag{3.5}$$

Maximizing (5) with respect to the parameters of interest delivers estimates for $\hat{\beta} = [1,\ \hat{\beta}_1,\ \hat{\beta}_2]$ and $\hat{\sigma}$ with an estimate of beliefs recovered as $\tilde{p}_{green}(f) = \hat{\beta}_1/\hat{\beta}_2$ and standard error recovered via the delta method. We alter this basic framework in one dimension: we restrict the estimate of $\tilde{p}_{green}(f)$ to be in the interval $[0, 1]$ by estimating a belief parameter, a, such that $\tilde{p}_{green}(f) = 1/(1 + exp(a))$. This change of variables allows us to investigate when beliefs alone are insufficient to rationalize key observations.

Note that this framework can be extended to multiple ambiguous acts. Beliefs induced by act g, $\tilde{p}_{red}(g)$, can be estimated alongside $\tilde{p}_{green}(f)$ by altering (3) to

$$\begin{aligned} q_j =& 1 + \mathbf{1}_{fj}\alpha_j \cdot \tilde{p}_{green}(f) \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]} \\ &+ \mathbf{1}_{gj}\alpha_j \cdot \tilde{p}_{red}(g) \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]} + (1 - \alpha_j) \cdot p_{cj} \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]} + u_j, \end{aligned} \tag{3.6}$$

where $\mathbf{1}_{fj}$ and $\mathbf{1}_{gj}$ indicate acts f and g being under consideration in choice j, respectively. Note that for act f, $\tilde{p}_{green}(f)$ summarized the belief that the act would yield the low outcome $x < y$, corresponding to the draw of green marble. Similarly for act g, $\tilde{p}_{red}(g)$ summarizes the belief that the act yields the low outcome $x < y$, corresponding to the draw of red marble. The likelihood function (5) can be adjusted with similar ability to recover the key parameters of interest: risk aversion, $\tilde{p}_{green}(f)$, and $\tilde{p}_{red}(g)$. The SEU restriction of $\tilde{p}_{green}(f) + \tilde{p}_{red}(g) = 1$ can then be tested.

Table 3.A2 provides maximum likelihood estimates of decision-making parameters building from this foundation. Column (1) presents the estimates of (6) corresponding to SEU with the restriction $\tilde{p}_{green}(f) + \tilde{p}_{red}(g) = 1$. We estimate $\tilde{p}_{green}(f) = 0.495$ (0.009), $\tilde{p}_{red}(g) = 0.505$ (0.009) and a risk aversion parameter of -0.466 (0.021).[51] Column (2) allows for $\tilde{p}_{green}(f)$ and $\tilde{p}_{red}(g)$ to be estimated

[51] A risk neutral individual would have a parameter value of -0.666, lower numbers indicate risk

independently finding $\tilde{p}_{green}(f) = 0.888(0.031)$, $\tilde{p}_{red}(g) = 0.901(0.030)$. The beliefs that rationalize the data are that subjective acts deliver around 90% low-paying balls and 10% high-paying balls regardless of the labeling of the states. Consistent with our finding of subjective-subjective inconsistency, one rejects the SEU null hypothesis of $\tilde{p}_{green}(f) + \tilde{p}_{red}(g) = 1$, $(\chi^2(1) = 194.8,\ p < 0.01)$. Correspondingly, the beliefs for the good outcome, y, across subjective acts are substantially sub-additive, with an estimate of $\tilde{p}_{red}(f) + \tilde{p}_{green}(g) = 0.21$, (clustered s.e = 0.06).[52]

Column (3) of Table 3.A2 allows for the estimated beliefs $\tilde{p}_{green}(f)$ and $\tilde{p}_{red}(g)$ to depend upon whether the subjective act is mixed with h, with l, or with $\frac{1}{2}h + \frac{1}{2}l$.[53] Consistent with our aggregate findings on subjective-objective inconsistency, estimates show that when mixing with h individuals are substantially more pessimistic about subjective acts than when mixing with l. Indeed, when mixing with h, the estimate of beliefs is censored at 100% low-paying balls, indicating potentially that beliefs alone are insufficient to rationalize the high mixture data. The null hypothesis of consistent beliefs across mixtures is rejected, $(\chi^2(6) = 422.6,\ p < 0.01)$.[54] Additionally, allowing for differential treatment of mixtures, the null hypothesis $\tilde{p}_{green}(f) + \tilde{p}_{red}(g) = 1$ continues to be rejected based upon the fully subjective data alone, $(\chi^2(1) = 221.3,\ p < 0.01)$.

In the final column of Table 3.A2 we include the insights from the individual level dominance violations and allow for explicit subjective act consideration costs, K_f and K_g.[55] We estimate K_f and K_g around -0.04 and reject the null hypothesis

aversion as they imply the distance between the utility of \$10 and the utility of \$30 to be less than two-thirds the distance between the utility of \$0 and the utility of \$30.

[52] As these are the complementary probabilities of those estimated the test statistic for the null hypothesis that $\tilde{p}_{red}(f) + \tilde{p}_{green}(g) = 1$ is again $\chi^2(1) = 194.8,\ p < 0.01$.

[53] That is, we assume the functional form $a = \gamma_0 + \gamma_1 \mathbf{1}_h + \gamma_2 \mathbf{1}_l + \gamma_3 \mathbf{1}_l \mathbf{1}_h$. Hence, $\tilde{p}_{green}(f \times l) = 1/(1 + exp(\gamma_0 + \gamma_2))$, $\tilde{p}_{green}(f \times h) = 1/(1 + exp(\gamma_0 + \gamma_1))$, and $\tilde{p}_{green}(f \times [\frac{1}{2}h + \frac{1}{2}l]) = 1/(1 + exp(\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3))$.

[54] Due to censored belief estimates, this hypothesis test conducted prior to transformation from belief parameter a to beliefs $\tilde{p}$. Null hypothesis is $\gamma_1 = \gamma_2 = \gamma_3 = 0$ for both act f and act g. A Likelihood Ratio test of column (3) vs. column (2) yields an identical conclusion, $\chi^2(6) = 72.03,\ p < 0.01$.

[55] The likelihood is augmented to reflect the decision cost of evaluating act f and act g, K_f and K_g. We assume K_f and K_g are constant over the j questions and are independent of α_j and the constant act c. We normalize consideration costs associated with considering the constant act c such that when $\alpha_j = 0$, consideration costs are zero. The new estimation equation is

of zero consideration costs, ($\chi^2(2) = 29.2,\ p < 0.01$).[56] The estimates indicate that considering a subjective act costs an individual around a 4% of winning $30 relative to an act with identical objective probability. Hence, not having to think about ambiguity is worth around a 4% chance of $30.

This appendix demonstrates the possibility of estimating decision-making parameters linked to our observed inconsistencies. We allow for three key deviations from subjective expected utility: subjective-subjective inconsistency, subjective-objective inconsistency, and explicit subjective act consideration costs. Taken together these effects are significant in the data with substantial improvements in fit at each stage. To demonstrate the fit of the final model, Figure 3.A4 relates predicted and actual values for Table 3.A2, column (4). Open circles correspond to predicted valuations and are presented for every point. The model is able to match key observations quite closely such as the behavior when $\alpha = 1$, and the broad patterns of behavior when mixing with high and low outcomes. Some systematic over-prediction exists for choices that are only risky potentially indicating some remaining misspecification. Broadly, however, the relationship between predicted and actual valuations is tightly grouped around the 45 degree line. The correlation between predicted and actual mean valuations is 0.96, though when regressing actual valuations on predicted valuations, one rejects the null hypothesis that the intercept is zero and the slope is 1, $F(2, 18) = 18.86,\ p < 0.01$. This suggests some remaining elements of choice not accounted for even in the rich model of column (4).

---

then

$$\begin{aligned} q_j =& 1 + \mathbf{1}_{fj} K_f + \mathbf{1}_{gj} K_g + \mathbf{1}_{fj} \alpha_j \cdot \tilde{p}_{green}(f) \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]} \\ &+ \mathbf{1}_{gj} \alpha_j \cdot \tilde{p}_{red}(g) \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]} + (1 - \alpha_j) \cdot p_{cj} \cdot \frac{[u(x) - u(y)]}{[u(y) - u(0)]} + u_j. \end{aligned} \tag{3.7}$$

Note that K_f is measured on the same scale as q_j such that the following sentence has empirical content: "Not thinking about ambiguity is worth a 5% chance of receiving $y." This as well can be adjusted to multiple subjective acts f and g with corresponding decision costs K_f and K_g.

56 The null hypothesis of zero consideration costs is also rejected by a likelihood ratio test between columns (3) and (4), $\chi^2(2) = 12.27,\ p < 0.01$.

Table 3.A2: Estimates of Decisionmaking Parameters

	(1)	(2)	(3)	(4)
$\tilde{p}_{green}(f)$	0.495	0.888	0.765	0.692
	(0.009)	(0.031)	(0.020)	(0.035)
$\tilde{p}_{red}(f) = (1 - \tilde{p}_{green}(f))$	0.505	0.112	0.235	0.308
	(0.009)	(0.031)	(0.020)	(0.035)
$\tilde{p}_{green}(f \times h)$			1	0.880
			(.)	(0.046)
$\tilde{p}_{green}(f \times l)$			0.659	0.537
			(0.026)	(0.052)
$\tilde{p}_{green}(f \times [\frac{1}{2}l + \frac{1}{2}h])$			0.831	0.676
			(0.036)	(0.064)
$\tilde{p}_{red}(g)$	0.505	0.901	0.763	0.695
	(0.009)	(0.030)	(0.023)	(0.037)
$\tilde{p}_{green}(g) = (1 - \tilde{p}_{red}(g))$	0.495	0.099	0.237	0.305
	(0.009)	(0.030)	(0.023)	(0.037)
$\tilde{p}_{red}(g \times h)$			1	0.898
			(.)	(0.052)
$\tilde{p}_{red}(g \times l)$			0.684	0.572
			(0.026)	(0.047)
$\tilde{p}_{red}(g \times [\frac{1}{2}l + \frac{1}{2}h])$			0.863	0.719
			(0.035)	(0.065)
$\frac{u(10)-u(30)}{u(30)-u(0)}$	-0.466	-0.369	-0.395	-0.376
	(0.021)	(0.017)	(0.017)	(0.017)
K_f				-0.042
				(0.010)
K_g				-0.040
				(0.009)
σ	0.168	0.158	0.155	0.155
	(0.007)	(0.007)	(0.007)	(0.007)
# Observations	2220	2220	2220	2220
# Clusters	111	111	111	111
Log-Likelihood	-5843.20	-5710.96	-5674.94	-5668.80
$H_0 : \tilde{p}_{green}(f) + \tilde{p}_{red}(g) = 1$		$\chi^2(1) = 194.8$	$\chi^2(1) = 221.3$	$\chi^2(1) = 41.1$
		$(p < 0.01)$	$(p < 0.01)$	$(p < 0.01)$
$H_0 : \tilde{p}_{red}(g) = \tilde{p}_{red}(g \times h) = \tilde{p}_{red}(g \times l) = \tilde{p}_{red}(g \times [\frac{1}{2}l + \frac{1}{2}h])$;			$\chi^2(6) = 422.6$	$\chi^2(6) = 109.4$
$\tilde{p}_{green}(f) = \tilde{p}_{green}(f \times h) = \tilde{p}_{green}(f \times l) = \tilde{p}_{green}(f \times [\frac{1}{2}l + \frac{1}{2}h])$			$(p < 0.01)$	$(p < 0.01)$
$H_0 : K_f = K_g = 0$				$\chi^2(2) = 29.2$
				$(p < 0.01)$

Maximum Likelihood Estimates with standard errors clustered on the individual level in parentheses. Belief estimates restricted to be in the interval $[0, 1]$ by estimating a belief parameter a such the $\tilde{p} = 1/(1 + exp(a))$. Hypothesis test of equal beliefs across mixtures conducted prior to transformation due to censored estimate in column (3). Conducted tests before and after transformation yield identical statistical conclusion as do Likelihood Ratio tests between columns.

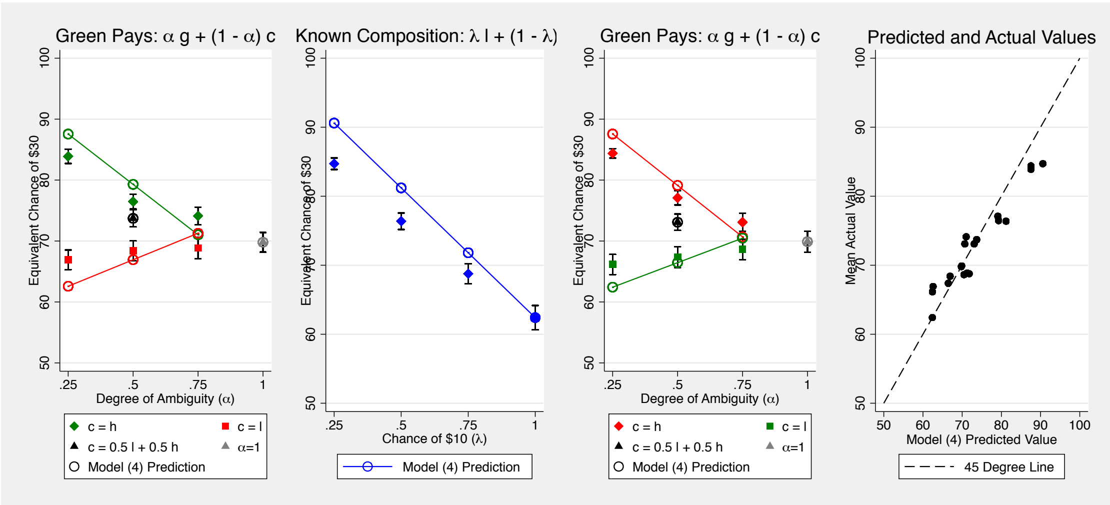


Figure 3.A4: Estimated mean valuations and standard errors from interval regressions (Stewart, 1983) for all experimental choices. See Appendix Table 3.A3 for further detail. Model prediction from Column (4) of Table 3.A2.

Supplementary Material

S.2 The Standard Portfolio Choice Problem in Germany

S.2.1 Instructions – SOEP study (original German)

Einwilligung zur Teilnahme

Im Folgenden bitten wir Sie, an einem “Finanzentscheidungsexperiment” teilzunehmen.

Sie können auf keinen Fall Geld verlieren!

Abhängig von Ihrer Entscheidung und zufälligen Faktoren, bekommen Sie am Ende der Befragung einen Geldbetrag tatsächlich ausbezahlt.

O Finanzentscheidungsexperiment starten
O Möchte nicht teilnehmen

Einwilligung zur Teilnahme – Nachfrage

Das “Finanzentscheidungsexperiment” ist Teil der Befragung, bei dem Sie zusätzlich einen Geldbetrag ausbezahlt bekommen. Sind Sie sicher, dass Sie nicht teilnehmen wollen?

O Finanzentscheidungsexperiment doch starten
O Möchte nicht teilnehmen, weil: . . . .

Baseline – Schirm 1

Wir bieten Ihnen eine Investitionsmöglichkeit an.

Stellen Sie sich bitte vor, 50.000 EUR aus eigenem Besitz zu investieren.

Diesen Betrag können Sie auf die folgenden beiden Geldanlagen verteilen:

1. Ein vom deutschen Staat ausgegebenes Wertpapier, das Ihnen einen Zins von 4% garantiert. Das Wertpapier wird im weiteren Text "Bundesanleihe" genannt.
2. Ein Bündel von Aktien, das im weiteren Text "Fonds" genannt wird. Der Gewinn oder Verlust dieses Fonds orientiert sich am Deutschen Aktien Index DAX, der die Entwicklung von 30 deutschen Großunternehmen zusammenfasst.

Wir werden Sie entsprechend Ihrer Entscheidung in einem kleineren Maßstab tatsächlich bezahlen.

Nehmen Sie sich Zeit, die Anweisungen in Ruhe durchzulesen und über Ihre Entscheidung nachzudenken.

Treatment – Schirm 1

Wir bieten Ihnen eine Investitionsmöglichkeit an.

Stellen Sie sich bitte vor, 50.000 EUR aus eigenem Besitz zu investieren.

Diesen Betrag können Sie auf die folgenden beiden Geldanlagen verteilen:

1. Ein vom deutschen Staat ausgegebenes Wertpapier, das Ihnen einen Zins von 4% garantiert. Das Wertpapier wird im weiteren Text "Bundesanleihe" genannt.
2. Ein Bündel von Aktien, das im weiteren Text "Fonds" genannt wird. Der Gewinn oder Verlust dieses Fonds orientiert sich am Deutschen Aktien Index DAX, der die Entwicklung von 30 deutschen Großunternehmen zusammenfasst.

Der Fonds schneidet entweder 5 Prozentpunkte besser oder 5 Prozentpunkte schlechter ab als der DAX. Welche der beiden Möglichkeiten zutreffen wird, erfahren Sie gleich.

Wir werden Sie entsprechend Ihrer Entscheidung in einem kleineren Maßstab tatsächlich bezahlen.

Nehmen Sie sich Zeit, die Anweisungen in Ruhe durchzulesen und über Ihre Entscheidung nachzudenken.

Baseline – Schirm 2

Sie verteilen zunächst, wie oben beschrieben, die 50.000 EUR auf Bundesanleihe und Fonds. Wir berechnen dann den Ertrag, den diese Investition erzielt.

- Für Geld, das Sie in die Bundesanleihe investieren, ist diese Berechnung einfach: Bei einem Zins von 4% machen Sie für jede 100 EUR, die Sie investieren einen sicheren Gewinn von 4 EUR.
- Um Gewinne und Verluste für Investitionen in den Fonds festzustellen, benutzen wir historische DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010. Der Computer wählt zufällig ein Jahr aus diesem Zeitraum aus und berechnet für dieses Jahr, was aus dem von Ihnen investierten Betrag geworden wäre.

Hier sehen Sie zwei Beispiele, die natürlich nur willkürlich sind und nichts über die tatsächliche Entwicklung des DAX aussagen:

Wenn der DAX in dem zufällig ausgewählten Jahr

- einen Gewinn von +15% erzielt hat, dann machen Sie für jede 100 EUR, die Sie in den Fonds investiert haben einen Gewinn von 15 EUR
- einen Verlust von -15% erzielt hat, dann verlieren Sie für jede 100 EUR, die Sie in den Fonds investiert haben, 15 EUR.

Ihr Gesamtgewinn ist dann einfach die Summe des Gewinns, den Sie durch Investitionen in die Bundesanleihe und den Fonds erzielen. Diesen Betrag zahlen wir Ihnen in kleinerem Maßstab aus. Für je 2000 EUR bekommen Sie am Ende des Experiments 1 EUR in bar ausbezahlt.

Treatment (minus) – Schirm 2

Sie verteilen zunächst, wie oben beschrieben, die 50.000 EUR auf Bundesanleihe und Fonds. Wir berechnen dann den Ertrag, den diese Investition erzielt.

- Für Geld, das Sie in die Bundesanleihe investieren, ist diese Berechnung einfach: Bei einem Zins von 4% machen Sie für jede 100 EUR, die Sie investieren einen sicheren Gewinn von 4 EUR.
- Um Gewinne und Verluste für Investitionen in den Fonds festzustellen, benutzen wir historische DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010. Der Computer wählt zufällig ein Jahr aus diesem Zeitraum aus und berechnet für dieses Jahr, was aus dem von Ihnen investierten Betrag geworden wäre.
  Zusätzlich wurde vom Computer zufällig bestimmt, dass Sie 5 Prozentpunkte weniger erhalten.

Hier sehen Sie drei Beispiele, die natürlich nur willkürlich sind und nichts über die tatsächliche Entwicklung des DAX aussagen: Wenn der DAX in dem zufällig ausgewählten Jahr

- einen Gewinn von +15% erzielt hat, dann macht der Fonds einen Gewinn von 15% - 5% = 10%. Das heißt, Sie machen für jede 100 EUR, die Sie in den Fonds investiert haben einen Gewinn von 10 EUR
- einen Verlust von -15% erzielt hat, dann macht der Fonds einen Verlust von -15% - 5% = -20%. Das heißt, Sie machen für jede 100 EUR, die Sie in den Fonds investiert haben einen Verlust von -20 EUR.
- Einen Gewinn von +2% gemacht hat, dann macht der Fonds einen Verlust von 2% - 5% = -3%. Das heißt, Sie machen für jede 100 EUR, die Sie in den Fonds investiert haben, einen Verlust von -3 EUR.

Ihr Gesamtgewinn ist dann einfach die Summe des Gewinns, den Sie durch Investitionen in die Bundesanleihe und den Fonds erzielen. Diesen Betrag zahlen wir Ihnen in kleinerem Maßstab aus. Für je 2000 EUR bekommen Sie am Ende des Experiments 1 EUR in bar ausbezahlt.

Treatment (plus) – Schirm 2

Sie verteilen zunächst, wie oben beschrieben, die 50.000 EUR auf Bundesanleihe und Fonds. Wir berechnen dann den Ertrag, den diese Investition erzielt.

- Für Geld, das Sie in die Bundesanleihe investieren, ist diese Berechnung einfach: Bei einem Zins von 4% machen Sie für jede 100 EUR, die Sie investieren einen sicheren Gewinn von 4 EUR.
- Um Gewinne und Verluste für Investitionen in den Fonds festzustellen, benutzen wir historische DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010. Der Computer wählt zufällig ein Jahr aus diesem Zeitraum aus und berechnet für dieses Jahr, was aus dem von Ihnen investierten Betrag geworden wäre.
  Zusätzlich wurde vom Computer zufällig bestimmt, dass Sie 5 Prozentpunkte mehr erhalten.

Hier sehen Sie drei Beispiele, die natürlich nur willkürlich sind und nichts über die tatsächliche Entwicklung des DAX aussagen:

Wenn der DAX in dem zufällig ausgewählten Jahr

- einen Gewinn von +15% erzielt hat, dann macht der Fonds einen Gewinn von 15% + 5% = 20%. Das heißt, Sie machen für jede 100 EUR, die Sie in den Fonds investiert haben einen Gewinn von 20 EUR
- einen Verlust von -15% erzielt hat, dann macht der Fonds einen Verlust von -15% + 5% = -10%. Das heißt, Sie machen für jede 100 EUR, die Sie in den Fonds investiert haben einen Verlust von -10 EUR.
- einen Verlust von -2% erzielt hat, dann macht der Fonds einen Gewinn von -2% + 5% = 3%. Das heißt, Sie machen für jede 100 EUR, die Sie in den Fonds investiert haben, einen Gewinn von 3 EUR.

Ihr Gesamtgewinn ist dann einfach die Summe des Gewinns, den Sie durch Investitionen in die Bundesanleihe und den Fonds erzielen. Diesen Betrag zahlen wir Ihnen in kleinerem Maßstab aus. Für jede 2000 EUR bekommen Sie am Ende des Experiments 1 EUR in bar ausbezahlt.

Baseline – Schirm 3

Zusammenfassend: Die Bundesanleihe wirft also in jedem Fall eine Verzinsung von 4% ab, während der Fonds für Ihre Auszahlung jeden der DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010 erzielen kann.

Wie viel der 50.000 EUR investieren Sie in die Bundesanleihe und wie viel in den Fonds?

Bitte achten Sie darauf, dass die beiden Beträge zusammen genau 50.000 EUR ergeben.

In die Bundesanleihe	.........Euro
In den Fonds	.........Euro

Treatment (minus) – Schirm 3

Zusammenfassend: Die Bundesanleihe wirft also in jedem Fall eine Verzinsung von 4% ab, während der Fonds für Ihre Auszahlung jeden der DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010, abzüglich der 5 Prozentpunkte, erzielen kann.

Wie viel der 50.000 EUR investieren Sie in die Bundesanleihe und wie viel in den Fonds?

Bitte achten Sie darauf, dass die beiden Beträge zusammen genau 50.000 EUR ergeben.

In die Bundesanleihe	.........Euro
In den Fonds	.........Euro

Treatment (plus) – Schirm 3

Zusammenfassend: Die Bundesanleihe wirft also in jedem Fall eine Verzinsung von 4% ab, während der Fonds für Ihre Auszahlung jeden der DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010, zuzüglich der 5 Prozentpunkte, erzielen kann.

Wie viel der 50.000 EUR investieren Sie in die Bundesanleihe und wie viel in den Fonds?

Bitte achten Sie darauf, dass die beiden Beträge zusammen genau 50.000 EUR ergeben.

In die Bundesanleihe	.........Euro
In den Fonds	.........Euro

Baseline – Schirm 4

Wie Sie wissen, hängt die Entwicklung des Fonds von der Entwicklung des DAX in den Jahren 1951 bis 2010 ab.

Im Folgenden wollen wir Sie fragen, wie Sie die möglichen Zahlungen des Fonds einschätzen.

Hierfür fassen wir auf dem nächsten Bildschirm die möglichen Verluste und Gewinne des Fonds in den folgenden sieben Bereichen zusammen:

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Über den sieben Bereichen befinden sich auf dem nächsten Schirm je 20 Kästchen. Zeigen Sie uns für diese sieben Bereiche an, wie häufig Sie den Fonds im jeweiligen Bereich vermuten, indem Sie die Kästen über den sieben Bereichen anklicken. Markieren Sie genau 20 Kästchen. Ein Kästchen steht für eine Häufigkeit von 1 zu 20, also 5 Prozent.

Durch das Markieren der Kästchen zeigen Sie uns, für wie wahrscheinlich Sie es halten, dass Ihr Fonds einen Verlust bzw. Gewinn in dem entsprechenden Bereich erzielt.

- Markieren Sie beispielsweise in einem Bereich gar keine Kästchen, so bringen Sie damit zum Ausdruck, dass Sie sich sicher sind, dass der Verlust oder Gewinn Ihres Fonds nicht in diesem Bereich liegt.
- Markieren Sie ein oder zwei Kästchen in einem Bereich, so halten Sie einen Verlust oder Gewinn in diesem Bereich für möglich aber nicht sehr wahrscheinlich
- Mehr Kästchen — bis zu 20 in einem Bereich — stehen für entsprechend höhere Wahrscheinlichkeiten.

Treatment (minus) – Schirm 4

Wie Sie wissen, hängt die Entwicklung des Fonds von der Entwicklung des DAX in den Jahren 1951 bis 2010 ab. Der Fonds liegt dabei immer 5 Prozentpunkte unter dem, was der DAX in einem dieser Jahre gezahlt hätte.

Im Folgenden wollen wir Sie fragen, wie Sie die möglichen Zahlungen des Fonds einschätzen.

Hierfür fassen wir auf dem nächsten Bildschirm die möglichen Verluste und Gewinne des Fonds in den folgenden sieben Bereichen zusammen:

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Über den sieben Bereichen befinden sich auf dem nächsten Schirm je 20 Kästchen. Zeigen Sie uns für diese sieben Bereiche an, wie häufig Sie den Fonds im jeweiligen Bereich vermuten, indem Sie die Kästen über den sieben Bereichen anklicken. Markieren Sie genau 20 Kästchen. Ein Kästchen steht für eine Häufigkeit von 1 zu 20, also 5 Prozent.

Durch das Markieren der Kästchen zeigen Sie uns, für wie wahrscheinlich Sie es halten, dass Ihr Fonds einen Verlust bzw. Gewinn in dem entsprechenden Bereich erzielt.

- Markieren Sie beispielsweise in einem Bereich gar keine Kästchen, so bringen Sie damit zum Ausdruck, dass Sie sich sicher sind, dass der Verlust oder Gewinn Ihres Fonds nicht in diesem Bereich liegt.
- Markieren Sie ein oder zwei Kästchen in einem Bereich, so halten Sie einen Verlust oder Gewinn in diesem Bereich für möglich aber nicht sehr wahrscheinlich
- Mehr Kästchen — bis zu 20 in einem Bereich — stehen für entsprechend höhere Wahrscheinlichkeiten.

Treatment (plus) – Schirm 4

Wie Sie wissen, hängt die Entwicklung des Fonds von der Entwicklung des DAX in den Jahren 1951 bis 2010 ab. Der Fonds liegt dabei immer 5 Prozentpunkte über dem, was der DAX in einem dieser Jahre gezahlt hätte.

Im Folgenden wollen wir Sie fragen, wie Sie die möglichen Zahlungen des Fonds einschätzen.

Hierfür fassen wir auf dem nächsten Bildschirm die möglichen Verluste und Gewinne des Fonds in den folgenden sieben Bereichen zusammen:

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Über den sieben Bereichen befinden sich auf dem nächsten Schirm je 20 Kästchen. Zeigen Sie uns für diese sieben Bereiche an, wie häufig Sie den Fonds im jeweiligen Bereich vermuten, indem Sie die Kästen über den sieben Bereichen anklicken. Markieren Sie genau 20 Kästchen. Ein Kästchen steht für eine Häufigkeit von 1 zu 20, also 5 Prozent.

Durch das Markieren der Kästchen zeigen Sie uns, für wie wahrscheinlich Sie es halten, dass Ihr Fonds einen Verlust bzw. Gewinn in dem entsprechenden Bereich erzielt.

- Markieren Sie beispielsweise in einem Bereich gar keine Kästchen, so bringen Sie damit zum Ausdruck, dass Sie sich sicher sind, dass der Verlust oder Gewinn Ihres Fonds nicht in diesem Bereich liegt.
- Markieren Sie ein oder zwei Kästchen in einem Bereich, so halten Sie einen Verlust oder Gewinn in diesem Bereich für möglich aber nicht sehr wahrscheinlich
- Mehr Kästchen – bis zu 20 in einem Bereich – stehen für entsprechend höhere Wahrscheinlichkeiten.

Baseline – Schirm 5

Markieren Sie jetzt bitte die 20 Kästchen so, dass Sie Ihre Einschätzung der Wertveränderung des Fonds widerspiegeln. Beachten Sie dabei alle für Sie denkbaren Möglichkeiten, die sich aus der historischen DAX-Entwicklung ergeben.

Sollten Sie zu diesem Zeitpunkt Ihre Investitionsentscheidung noch einmal ändern wollen, drücken Sie bitte auf "Zurück".

Füllen Sie die Kästchen immer, ohne Lücken, von UNTEN nach OBEN auf!

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Treatment (plus & minus) – Schirm 5

Markieren Sie jetzt bitte die 20 Kästchen so, dass Sie Ihre Einschätzung der Wertveränderung des Fonds widerspiegeln. Beachten Sie dabei alle für Sie denkbaren Möglichkeiten, die sich aus der historischen DAX-Entwicklung und dem (Aufschlag/Abschlag) von 5 Prozentpunkten ergeben.

Sollten Sie zu diesem Zeitpunkt Ihre Investitionsentscheidung noch einmal ändern wollen, drücken Sie bitte auf "Zurück".

Füllen Sie die Kästchen immer, ohne Lücken, von UNTEN nach OBEN auf!

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Baseline & Treatment – Schirm 6

Geben Sie bitte außerdem an, welche durchschnittliche Wertveränderung (in %) Sie für den Fonds erwarten.

→ *Bitte maximal auf eine Stelle nach dem Komma eingeben (z.B. xx.x)!*

→ *Bitte Punkt anstatt Komma eingeben*

Durchschnittliche Wertsteigerung ....
oder
Durchschnittlicher Wertverlust: ....

Baseline & Treatment – Schirm 7

Wir würden Ihnen nun gern ein paar Fragen zu dem soeben absolvierten Experiment stellen.

Wie Sie diese Fragen beantworten wird keinen Einfluss auf Ihre Auszahlung haben.

Wie sicher sind Sie sich Ihrer Einschätzung des Fonds?

Antworten Sie bitte anhand der folgenden Skala, bei der “0” gar nicht sicher und der Wert “10” sehr sicher bedeutet.

Mit den Werten zwischen “0” und “10” können Sie Ihre Meinung abstufen.

Gar nicht sicher O O O O O O O O O O O Sehr sicher

Baseline & Treatment – Schirm 8

Wahr oder falsch? Wenn der DAX in dem zufällig ausgewählten Jahr einen Gewinn von 40% gemacht hat, so wirft auch der Ihnen angebotene Fonds einen Gewinn von 40% ab.

O wahr
O falsch

Wahr oder falsch? Wenn der DAX in dem zufällig ausgewählten Jahr einen Verlust von 4% gemacht hat, so erzielt der Ihnen angebotete Fonds einen Verlust von -1%

O wahr
O falsch

Baseline & Treatment – Schirm 9

Nachdem es in den bisherigen Fragen um die Entwicklung eines an den DAX gekoppelten Fonds in der Vergangenheit ging wüssten wir nun gern, was Sie für die zukünftige Entwicklung des DAX selbst erwarten. Geben Sie auf dem nächsten Bildschirm an, wo Sie den DAX in einem Jahr sehen, ausgedrückt in Gewinn oder Verlust gegenüber dem heutigen Wert. Wir fassen dazu erneut die möglichen Gewinne und Verluste in die sieben größeren Bereiche zusammen.

Wir bitten Sie auch hier, alle für Sie denkbaren Entwicklungen des DAX in Betracht zu ziehen.

Zeigen Sie uns dann an, für wie wahrscheinlich Sie die jeweiligen Gewinne und Verluste halten.

Bitte drücken Sie dies aus, indem Sie wieder die 20 Kästchen markieren.

Ein Kästchen steht hier wieder für eine Häufigkeit von 1 zu 20, also 5 Prozent.

Durch das Markieren der Kästchen zeigen Sie uns für wie wahrscheinlich Sie die Wertveränderung des DAX, in einem Jahr, in einem der sieben Bereiche halten.

- Markieren Sie beispielsweise in einem Bereich gar keine Kästchen, so bringen Sie damit zum Ausdruck, dass Sie sich sicher sind, dass die Wertveränderung des DAX nicht in diesem Bereich liegt.
- Markieren Sie ein oder zwei Kästchen in einem Bereich, so halten Sie die Wertveränderung des DAX in diesem Bereich für möglich aber nicht sehr wahrscheinlich.
- Mehr Kästchen – bis zu 20 in einem Bereich – stehen für entsprechend höhere Wahrscheinlichkeiten.

Baseline & Treatment – Schirm 10

Markieren Sie jetzt bitte die 20 Kästchen so, dass Sie Ihre Einschätzung der DAX-Gewinne und DAX-Verlust in den nächsten 12 Monaten, also bis zum 19.11.2013 widerspiegeln.

Füllen Sie die Kästchen immer, ohne Lücken, von UNTEN nach OBEN auf!

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Baseline & Treatment – Schirm 11

Außerdem interessiert uns auch hier, wie sicher Sie sich Ihrer Einschätzung des DAX sind.

Wie sicher sind Sie sich Ihrer Einschätzung des DAX?

Antworten Sie bitte anhand der folgenden Skala, bei der “0” gar nicht sicher und der Wert “10” sehr sicher bedeutet.

Mit den Werten zwischen “0” und “10” können Sie Ihre Meinung abstufen.

Gar nicht sicher O O O O O O O O O O O Sehr sicher

Baseline & Treatment – Auszahlungsübersicht

Der Computer hat per Zufall das Jahr 1975 ausgewählt.

In diesem Jahr hat der DAX einen Gewinn von 41.21%,

und der Fonds somit einen Gewinn von 36.21% gemacht.

Wir zahlen Ihnen deshalb auf Basis Ihrer Investition 31 EUR aus, die sich wie folgt berechnen:

Anlage	Investition	Gewinn/Verlust	Auszahlung
Bundesanleihe	20000 EUR	4,0%	20800 EUR
Fonds	30000 EUR	36,21%	40863 EUR
		Summe	61663 EUR

Auszahlung	30,83 EUR
Auf den nächsten Euro gerundet	31 EUR

Das Finanzentscheidungsexperiment ist nun zu Ende

⇒ *Der Auszahlungsbetrag wird am Ende des Interviews nochmal angezeigt!*

S.2.2 Instructions – SOEP study (English translation)

Agreement to Participate

In the following we kindly ask you to take part in a "financial decision experiment".

You cannot possibly lose any money!

Depending on the decisions you will make and some random factors you will, however, receive some actual money at the end of the survey.

O Start the financial decision experiment
O I do not want to participate

Agreement to Participate – Second Take

The "financial decision experiment" is a part of this survey in which you can earn some money. Are you sure that you do not want to participate?

O I have changed my mind: Start the financial decision experiment
O I do not want to participate because: ....

Baseline – Screen 1

We offer you an investment opportunity.

Please imagine that you would like to invest 50,000 EUR of your own savings.

You can distribute this amount between the following investments:

1. A German sovereign bond that guarantees you an interest rate of 4%. We will call this asset the "Bund" henceforth.
2. A bundle of stocks that will be called the "fund". The gains and losses on this fund will be based on the German stock market index DAX, which is a summary measure of the performance of 30 major German enterprises.

We will pay you according to your decision on a smaller scale.

Please take your time to carefully read the instructions and think about your decision.

Treatment – Screen 1

We offer you an investment opportunity.

Please imagine that you would like to invest 50,000 EUR of your own savings.

You can distribute this amount between the following investments:

1. A German sovereign bond that guarantees you an interest rate of 4%. We will call this asset the "Bund" henceforth.
2. A bundle of stocks that will be called the "fund". The gains and losses on this fund will be based on the German stock market index DAX, which is a summary measure of the performance of 30 major German enterprises.

The return of the fund will be either 5 percentage points higher or 5 percentage points lower than that of the DAX. You will find out which of these two possibilities applies to you soon.

We will pay you according to your decision on a smaller scale.

Please take your time to carefully read the instructions and think about your decision.

Baseline – Screen 2

Please distribute the 50,000 EUR over Bund and fund as described above. We will then calculate the total return on your investment.

- For money invested in the Bund the calculation is simple: For each 100 EUR you invest in the Bund at an interest rate of 4% you will make sure profit of 4 EUR.
- Gains and losses on investments in the fund will be based on historical DAX gains and losses from 1951 to 2010. The computer will randomly choose a year in this time period and calculate for this exact year how your investment would have fared.

The following two examples are arbitrary and do not say anything about the actual performance of the DAX:

If the DAX in the randomly chosen year had made

- a gain of +15%, you would have earned 15 EUR for each 100 EUR invested in fund.
- a loss of -15%, you would have lost 15 EUR for each 100 EUR invested in fund.

Your total profit will be the sum of the profits of your investments in both Bund and fund. We will actually pay you this amount on a smaller scale. At the end of the experiment you will receive 1 EUR in cash for each 2000 EUR.

Treatment (minus) – Screen 2

Please distribute the 50,000 EUR over Bund and fund as described above. We will then calculate the total return on your investment.

- For money invested in the Bund the calculation is simple: For each 100 EUR you invest in the Bund at an interest rate of 4% you will make sure profit of 4 EUR.
- Gains and losses on investments in the fund will be based on historical DAX gains and losses from 1951 to 2010. The computer will randomly choose a year in this time period and calculate for this exact year how your investment would have fared.
  Additionally the computer has determined through a random draw that you will receive 5 percentage points less.

The following two examples are arbitrary and do not say anything about the actual performance of the DAX:

If the DAX in the randomly chosen year had made

- a gain of +15%, the fund would make a gain of 15%-5%=10%. This means that for each 100 EUR invested in fund you would earn 10 EUR.
- a loss of -15%, the fund would make a loss of -15%-5%=-20%. This means that for each 100 EUR invested in fund you would lose 20 EUR.
- a gain of +2%, then the fund would make a loss of 2%-5%=-3%. This means that for each 100 EUR invested in fund you would lose 3 EUR.

Your total profit will be the sum of the profits of your investments in both Bund and fund. We will actually pay you this amount on a smaller scale. At the end of the experiment you will receive 1 EUR in cash for each 2000 EUR.

Treatment (plus) – Screen 2

Please distribute the 50,000 EUR over Bund and fund as described above. We will then calculate the total return on your investment.

- For money invested in the Bund the calculation is simple: For each 100 EUR you invest in the Bund at an interest rate of 4% you will make sure profit of 4 EUR.
- Gains and losses on investments in the fund will be based on historical DAX gains and losses from 1951 to 2010. The computer will randomly choose a year in this time period and calculate for this exact year how your investment would have fared.
  Additionally the computer has determined through a random draw that you will receive 5 percentage points more.

The following two examples are arbitrary and do not say anything about the actual performance of the DAX:

If the DAX of the randomly chosen year had made

- a gain of +15%, the fund would make a gain of 15%+5%=20%. This means that for each 100 EUR invested in fund you would earn 20 EUR.
- a loss of -15%, the fund would make a loss of -15%+5%=-10%. This means that for each 100 EUR invested in fund you would lose 10 EUR.
- a loss of -2%, the fund would make a gain of -2%+5%=3%. This means that for each 100 EUR invested in fund you would earn 3 EUR.

Your total profit will be the sum of the profits of your investments in both Bund and fund. We will actually pay you this amount on a smaller scale. At the end of the experiment you will receive 1 EUR in cash for each 2000 EUR.

Baseline – Screen 3

To sum up: The Bund guarantees you an interest of 4% while the fund can produce any of the DAX-gain or DAX-losses from the years 1951 to 2010.

How much of the 50,000 EUR do you want to invest in the Bund and how much do you want to invest in the fund?

Please make sure that the two amounts sum up to exactly 50,000.

Bund ....Euro
fund ....Euro

Treatment (minus) – Screen 3

To sum up: The Bund guarantees you an interest of 4% while the fund can produce any of the DAX-gain or DAX-losses from the years 1951 to 2010, minus the 5 percentage points.

How much of the 50,000 EUR do you want to invest in the Bund and how much do you want to invest in the fund?

Please make sure that the two amounts sum up to exactly 50,000.

Bund ....Euro
fund ....Euro

Treatment (plus) – Screen 3

To sum up: The Bund guarantees you an interest of 4% while the fund can produce any of the DAX-gain or DAX-losses from the years 1951 to 2010, plus the 5 percentage points.

How much of the 50,000 EUR do you want to invest in the Bund and how much do you want to invest in the fund?

Please make sure that the two amounts sum up to exactly 50,000.

Bund	.........Euro
fund	.........Euro

Baseline – Screen 4

As you know, the development of the fund depends on the development of the DAX from 1951 to 2010.

In the following we would like to ask you for your expectations of the fund's possible payoffs.

For this purpose we will group the possible gains and losses of the fund into seven ranges on the next screen.

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

On the next screen 20 boxes will be placed above each of the seven ranges. Please indicate for all seven ranges how often you expect the fund to be in each range by clicking on the mentioned boxes. Please mark exactly twenty boxes. One box stands for a frequency of 1 in 20, i.e. for 5 percent.

By marking the boxes you will show us how likely you believe it is that your fund will produce a gain or loss in the given range.

- If, for instance, you don't mark any of the boxes in a particular range, this will mean that you are sure that the gain or loss will never lie in this range.
- If you mark one or two boxes in a particular range, you believe a loss or gain in this range to be possible, but not very likely.
- More boxes -– up to 20 in one range – imply correspondingly higher probabilities.

Treatment (minus) – Screen 4

As you know, the development of the fund depends on the development of the DAX from 1951 to 2010. The fund will always be 5 percentage points below the outcome that the DAX would have payed in one of these years.

In the following we would like to ask you for your expectations of the fund's possible payoffs.

For this purpose we will group the possible gains and losses of the fund into seven ranges on the next screen.

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

On the next screen 20 boxes will be placed above each of the seven ranges. Please indicate for all seven ranges how often you expect the fund to be in each range by clicking on the mentioned boxes. Please mark exactly twenty boxes. One box stands for a frequency of 1 in 20, i.e. for 5 percent.

By marking the boxes you will show us how likely you believe it is that your fund will produce a gain or loss in the given range.

- If, for instance, you don't mark any of the boxes in a particular range, this will mean that you are sure that the gain or loss will never lie in this range.
- If you mark one or two boxes in a particular range, you believe a loss or gain in this range to be possible, but not very likely.
- More boxes -– up to 20 in one range – imply correspondingly higher probabilities.

Treatment (plus) – Screen 4

As you know, the development of the fund depends on the development of the DAX from 1951 to 2010. The fund will always be 5 percentage points above the outcome that the DAX would have payed in one of these years.

In the following we would like to ask you for your expectations of the fund's possible payoffs.

For this purpose we will group the possible gains and losses of the fund into seven ranges on the next screen.

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

On the next screen 20 boxes will be placed above each of the seven ranges. Please indicate for all seven ranges how often you expect the fund to be in each range by clicking on the mentioned boxes. Please mark exactly twenty boxes. One box stands for a frequency of 1 in 20, i.e. for 5 percent.

By marking the boxes you will show us how likely you believe it is that your fund will produce a gain or loss in the given range.

- If, for instance, you don't mark any of the boxes in a particular range, this will mean that you are sure that the gain or loss will never lie in this range.
- If you mark one or two boxes in a particular range, you believe a loss or gain in this range to be possible, but not very likely.
- More boxes -– up to 20 in one range – imply correspondingly higher probabilities.

Baseline – Screen 5

Please mark the 20 boxes such that they reflect your assessment of the development of the fund. Please consider every — in your opinion — possible historical DAX development.

If you would like to reconsider and change your investment decision, please click the button “Back”.

Mark the boxes, avoiding gaps from BOTTOM to TOP!

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

Treatment (plus & minus) – Screen 5

Please mark the 20 boxes such that they reflect your assessment of the development of the fund. Please consider every — in your opinion — possible combination of the historical DAX development and the (addition/deduction) of 5 percentage points.

If you would like to reconsider and change your investment decision, please click the button "Back".

Mark the boxes, avoiding gaps from BOTTOM to TOP!

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

Baseline & Treatment – Screen 6

Please also let us know what average return (*lit*: "change in value") (in %) you expect for the fund.

→ *Please use a maximum of one decimal! (e.g. xx.x)*

→ *Please use a decimal point instead of a comma*

Average increase in value ....
or
Average decrease in value: ....

Baseline & Treatment – Screen 7

We would like to ask you some questions about the experiment which you have just completed.

Your answers to these questions will not influence your payment.

How confident are you in your assessment of the fund?

Please answer according to the following scale, in which "0" means "not at all confident" and the value "10" means "very confident".

With the values between "0" and "10" you can grade your opinion.

Not at all confident O O O O O O O O O O O Very confident

Baseline & Treatment – Screen 8

True or false? If the DAX made a gain of 40% in the randomly chosen year, the fund you have been offered would also make a gain of 40%.

O true

O false

True or false? If the DAX made a loss of 4% in the randomly chosen year, the fund you have been offered would make a loss of -1%.

O true

O false

Baseline & Treatment – Screen 9

The questions so far all concerned the development of a fund whose returns were tied to the development of the DAX in the past. We would now like to ask you some questions concerning your expectations for the *future* development of the DAX itself. On the next screen, please let us know where you see the DAX in one year, expressed as a gain or loss relative to its current value. We will again group the possible gains and losses into seven larger ranges.

Again we ask you to consider all of the developments of the DAX that you believe are possible. Please indicate how likely you think the different profits and losses to be. Please express this by again marking 20 boxes. As before, one box stands for a frequency of 1 out of 20, i.e. 5 percent.

By marking the boxes you will show us how likely you consider the change in value of the DAX in one year to lie in each of the 7 ranges

- If, for instance, you don't mark any of the boxes in a particular range, this will mean that you are sure that the gain or loss will not lie in this range.
- If you mark one or two boxes in a particular range, you believe a loss or gain in this range to be possible, but not very likely.
- More boxes -— up to 20 in one range — imply correspondingly higher probabilities.

Baseline & Treatment – Screen 10

Please mark the 20 boxes according to your assessment of the development of the DAX-profits and DAX-losses in the next 12 months, i.e. until 19.11.2013.

Mark the boxes, avoiding gaps from BOTTOM to TOP!

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

Baseline & Treatment – Screen 11

Morevoer, we are interested in how sure you are about your assessment of the DAX.

How confident are you in your assessment of the DAX?

Please answer according to the following scale, in which “0” means “not at all confident” and the value “10” means “very confident”.

With the values between “0” and “10” you can grade your opinion.

Not at all confident O O O O O O O O O O O Very confident

Baseline & Treatment – Payout Overview

The Computer randomly chose the year 1975.

In this year the DAX incurred a profit of 41.21%

which means that fund incurred a profit of 36.21%

As a result, we will pay you 31 EUR based on your investment, according to the following calculation:

Asset	Invested Amount	Gain/Loss	Payment
Bundesanleihe	20.000 EUR	4.0%	20.800 EUR
Fonds	30.000 EUR	36.21%	40.863 EUR
		Sum	61.663 EUR

Payment	30.83 EUR
Rounded up to the next Euro	31 EUR

This concludes the financial decision experiment.

⇒ *The amount of payment will reappear on the screen at the end of the interview.*

S.2.3 Instructions – Complexity Study (original German)

Willkommens-Schirm

Willkommen!

Im Folgenden bitten wir Sie, an einem Finanzentscheidungsexperiment teilzunehmen.

Abhängig von Ihrer Entscheidung und zufälligen Faktoren, bekommen Sie am Ende der Befragung einen Geldbetrag tatsächlich ausbezahlt. Sie können dabei auf keinen Fall Geld verlieren.

Es ist wichtig, dass Sie während des Experiments still bleiben und nicht mit anderen Teilnehmern kommunizieren. Sollten Sie Fragen haben oder Hilfe brauchen, dann heben Sie bitte die Hand, und ein Experimentator wird zu Ihnen kommen. Sollten Sie sich nicht an diese Anweisung halten, so müssen wir Sie vom Experiment ausschließen. Vielen Dank.

Schirm 1

Im Folgenden müssen Sie in 8 Runden jeweils eine Investitionsentscheidung fällen. Alle Runden sind gleich aufgebaut. Eine der 8 Runden wird am Ende des Experiments zufällig ausgewählt und Ihnen tatsächlich ausbezahlt. Wie genau das passiert, dazu gleich gleich mehr.

Sie haben in jeder Runde jeweils eine Summe Geld zur Verfügung, die Sie zwischen zwei Geldanlagen aufteilen müssen. Außerdem bekommen Sie jeweils einen zusätzlichen festen Geldbetrag, unabhängig von Ihrer Entscheidung in dieser Runde.

Eine der beiden Geldanlagen hat einen festen Zinssatz. Die andere Geldanlage hat einen Zinssatz, der von der Entwicklung am Aktienmarkt abhängt. Darüber hinaus gibt es pro Runde auf jede der beiden Geldanlagen einen Bonus (das heißt der Zinssatz wird um einen festen Betrag erhöht).

Die Geldanlage mit dem festen Zinssatz zahlt in jeder Runde 2% Zinsertrag zuzüglich des Bonus. Diese Geldanlage wird im weiteren Text "Bundesanleihe" genannt. Die andere Geldanlage orientiert sich am Deutschen Aktien Index DAX, der die Entwicklung von 30 deutschen Großunternehmen zusammenfasst. Um die Verzinsung dieser Geldanlage festzustellen, benutzen wir historische DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010, und addieren den entsprechenden Bonus hinzu. Der Computer wählt zufällig ein Jahr aus diesem Zeitraum aus und berechnet für dieses Jahr, was aus dem von Ihnen investierten Betrag geworden wäre. Diese Geldanlage wird im weiteren Text "Fonds" genannt.

Schirm 2

Ein Beispiel könnte wie folgt aussehen. Sie haben 50.000 Euro, die Sie auf Bundesanleihe und Fonds aufteilen müssen und eine Auszahlung von 14.000 Euro, die Sie unabhängig von Ihrer Entscheidung bekommen. Auf den Zins der Bundesanleihe erhalten Sie einen Bonus von 3 Prozentpunkten, auf den Zins des Fonds erhalten

Sie ebenfalls einen Bonus von 3 Prozentpunkten.

Konkret bedeutet das in diesem Beispiel, dass Sie für den Betrag, den Sie in die Bundesanleihe investieren, einen Zins von 5% erhalten: die stets gleichen 2% zuzüglich des in dieser Runde relevanten Bonus von 3%. Sie machen also für jede 100 Euro, die Sie in die Bundesanleihe investiert haben, einen Gewinn von 5 Euro, und bekommen am Ende 105 Euro ausgezahlt. Die Verzinsung des in den Fonds investierten Betrags wird in diesem Beispiel wie folgt bestimmt: Sie ist die realisierte Kursentwicklung des DAX in einem zufällig gezogenen Jahr (aus 1951 bis 2010) plus der Bonus von 3

- Hat also der DAX zum Beispiel in dem zufällig gezogenen Jahr einen Gewinn von 3,5% gemacht, erhalten Sie auf den Betrag, den Sie in den Fonds investiert haben, eine Verzinsung von 6,5%. Sie machen also für jede 100 Euro, die Sie in den Fonds investiert haben, einen Gewinn von 6,50 Euro, und bekommen am Ende 106,50 Euro ausgezahlt.
- Hat der DAX dagegen im zufällig gezogenen Jahr einen Gewinn von 12% gemacht, erhalten Sie auf Ihren investierten Betrag eine Verzinsung von 15% (mit 115 Euro Auszahlung pro 100 Euro Investition).
- Hat der DAX im zufällig gezogenen Jahr einen Verlust von 12% gemacht, so erhalten Sie eine negative Verzinsung, die aber wegen dem Bonus um 3% geringer ist, also ein Verlust von 9%. In diesem Fall würden Sie für jede 100 Euro Investition eine Auszahlung von 91 Euro bekommen.

Ihre Gesamtauszahlung ergibt sich in diesem Beispiel als 14.000 Euro (die feste Auszahlung) plus die verzinste Investition in die Bundesanleihe plus die verzinste Investition in den Fonds.

Bitte beachten Sie, dass der Bonus auf die Bundesanleihe sich später vom Bonus auf den Fonds unterscheiden wird. Sie sind nur in diesem Beispiel gleich hoch gewählt.

(Das Beispiel ist natürlich willkürlich und sagt nichts über die tatsächliche Entwicklung des DAX oder über andere unbekannten Größen aus.)

Schirm 3

Die Gesamtauszahlung zahlen wir Ihnen für eine der 8 Runden im kleineren Maßstab aus. Das heißt, der Computer wählt am Ende des Experiments zufällig eine der 8 Runden aus. Dabei hat jede Runde die gleiche Wahrscheinlichkeit, ausgewählt zu werden. Diese Runde wird Ihnen in bar ausbezahlt.

Zusätzlich zieht der Computer ebenso zufällig und mit gleicher Wahrscheinlichkeit ein Jahr aus dem Zeitraum 1951 bis 2010. Der Gewinn oder Verlust des DAX in diesem Jahr wird dann herangezogen, um Ihre Auszahlung zu bestimmen.

Für je 5000 Euro, die Sie in der Runde als Gesamtauszahlung bekommen, erhalten Sie 1 Euro in bar.

Zusammenfassend: Die Bundesanleihe wirft also eine Verzinsung von 2% zuzüglich des entsprechenden Bonus ab, während der Fonds für Ihre Auszahlung jeden der DAX-Gewinne und DAX-Verluste der Jahre 1951 bis 2010 zuzüglich des entsprechenden Bonus erzielen kann.

Investitions-Schirm

Runde 1

Sie haben 50.000 Euro, die Sie zwischen der Bundesanleihe und dem Fonds aufteilen müssen. Unabhängig von Ihrer Entscheidung erhalten Sie zusätzlich einen Betrag von 17.550 Euro.

Für die Bundesanleihe gibt es einen Bonus von 2,80 Prozentpunkten.

Für den Fonds gibt es einen Bonus von 5,90 Prozentpunkten.

Bonus auf die Bundesanleihe	Bonus auf den Fonds	Investitionssumme	zusätzliche Zahlung
2,80 Prozentpunkte	5,90 Prozentpunkte	50.000 Euro	17.550 Euro

Wie viel der 50.000 EUR investieren Sie in die Bundesanleihe und wie viel in den Fonds?

Bitte achten Sie darauf, dass beide Beträge ganze Zahlen sind und zusammen genau 50.000 EUR ergeben.

In die Bundesanleihe	.........Euro
In den Fonds	.........Euro

Schirm 4

Wie Sie wissen, hängt die Verzinsung des Fonds von der Entwicklung des DAX in den Jahren 1951 bis 2010 ab. Im Folgenden wollen wir Sie fragen, wie Sie die Gewinne und Verluste des DAX in diesem Zeitraum einschätzen.

Hierfür fassen wir auf dem nächsten Bildschirm die möglichen Gewinne und Verluste in den folgenden sieben Bereichen zusammen:

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Über den sieben Bereichen befinden sich auf dem nächsten Schirm je 20 Kästchen. Zeigen Sie uns für diese sieben Bereiche an, wie häufig Sie den DAX im jeweiligen Bereich vermuten, indem Sie die Kästchen über den sieben Bereichen anklicken.

Markieren Sie genau 20 Kästchen. Ein Kästchen steht für eine Häufigkeit von 1 zu 20, also 5 Prozent.

Durch das Markieren der Kästchen zeigen Sie uns Einschätzung darüber, wie häufig der DAX in den Jahren 1951-2010 einen Verlust bzw. Gewinn in dem entsprechenden Bereich erzielt, an.

- Markieren Sie beispielsweise in einem Bereich gar kein Kästchen, so bringen Sie damit zum Ausdruck, dass Sie sich sicher sind, dass der Verlust oder Gewinn des DAX nie in diesem Bereich lag.
- Markieren Sie ein oder zwei Kästchen in einem Bereich, so halten Sie einen Verlust oder Gewinn in diesem Bereich für möglich aber nicht sehr wahrscheinlich.
- Mehr Kästchen - bis zu 20 in einem Bereich - stehen für entsprechend höhere Wahrscheinlichkeiten.

Schirm 5

Markieren Sie jetzt bitte die 20 Kästchen so, dass Sie Ihre Einschätzung der Wertveränderung des DAX im Zeitraum 1951 bis 2010 widerspiegeln.

Füllen Sie die Kästchen immer, ohne Lücken, von UNTEN nach OBEN auf!

Verlust zwischen 60% und 90%	Verlust zwischen 30% und 90%	Verlust zwischen 0% und 30%	Gewinn zwischen 0% und 30%	Gewinn zwischen 30% und 60%	Gewinn zwischen 60% und 90%	Gewinn zwischen 90% und 120%

Auszahlungsübersicht

Das war's. Vielen Dank für Ihre Teilnahme!

Der Computer hat per Zufall bestimmt, dass Ihnen ihre Investition aus Runde 6 ausbezahlt wird. In dieser Runde haben Sie 25.000 EUR in die Bundesanleihe und 25.000 EUR in den DAX investiert. Der Bonus auf den Zins der Bundesanleihe betrug in dieser Runde 3,00 Prozentpunkte, der Bonus auf den DAX betrug 6,05 Prozentpunkte.

Der Computer hat außerdem per Zufall das Jahr 1992 ausgewählt. In diesem Jahr hat der DAX einen Verlust von 0,66% gemacht.

Wir zahlen Ihnen deshalb auf Basis ihrer Investition 14 EUR aus, die sich wie folgt berechnen:

Anlage	investierter Betrag	Gewinn / Verlust	Bonus/Malus	Gewinn / Verlust insgesamt	Auszahlung
Bundesanleihe	25.000 EUR	2 %	3,00 %	5,00 %	26.250 EUR
Fonds	25.000 EUR	-0,66 %	6,05 %	5,39 %	26.348 EUR
				Zusätzliche Zahlung	15.800 EUR
				Summe	68.398 EUR

Summe / 5000	13,68 EUR
(auf den nächsten EUR gerundet)	14 EUR

Bitte bleiben Sie noch einen Moment sitzen. Sobald die große Mehrzahl der Teilnehmer das Experiment abgeschlossen hat, werden wir mit der Auszahlung beginnen.

S.2.4 Instructions – Complexity Study (English translation)

Welcome Screen

Welcome!

In the following, we kindly ask you to take part in a financial decision experiment.

Depending on your decision and some random factors, you will receive an amount of money for real at the end of the experiment. You cannot possibly lose any money.

It is important that you remain silent throughout the experiment and that you do not communicate with other participants. Should you have any questions or need any help, please raise your hand and an experimenter will come to you. If you do not follow these instructions, we will have to exclude you from the experiment. Thank you very much.

Screen 1

In the following you have to make one investment decision in each of 8 rounds. All rounds are constructed in the same way. At the end of the experiment one of the 8 rounds will be randomly selected and the money earned in this round will be your payment. We will tell you more about the exact way this works shortly.

In each round you will have a certain amount of money, which you must distribute among two financial assets. Furthermore, in each round you will receive an additional fixed amount of money that you will receive independent of what your investment decision is.

One of the two financial assets offers a fixed rate of interest. The other asset has an interest rate which depends on the development of the stock market. In addition there will be a bonus applied to both assets (i.e. the interest rate will be increased by a fixed amount).

The investment possibility with the fixed interest rate pays 2% plus the bonus in each round. We will call this asset the "Bund" in the following text. The other financial asset will be based on the German stock market index DAX, which is a summary measure of the performance of 30 major German enterprises. To determine the return on this investment, we use historical DAX-profits and DAX-losses from the years 1951 to 2010 and then add the bonus. The computer randomly chooses a year in this period and calculates for this exact year how your invested sum would have fared. We will call this financial asset the "fund".

Screen 2

An example could be as follows. Imagine that you have 50,000 Euros, which you must distribute over Bund and fund, as well as a payment of 14,000 Euros that you receive independent of your investment decision. You receive a 3 percentage points bonus on the interest of the Bund, and a 3 percentage points bonus on the interest of the fund.

Concretely for this example, that means that you would receive an interest of 5% for the amount invested in the Bund: the usual 2% plus the relevant bonus of 3% for this round. For each 100 Euros invested in the Bund, you earn 5 Euros, and are paid 105 euros at the end. The interest of the amount invested in the fund is calculated as follows: It will be the realized return of the DAX in one randomly chosen year (from 1951 to 2010) plus the bonus of 3%.

- If the DAX made a profit of 3.5% in the randomly chosen year, you would would get an interest rate of 6.5 % on the amount invested in the fund. This means that you earn 6.50 Euros for each 100 Euros invested in the fund and are paid 106.50 Euros at the end of the experiment.
- If, in contrast, the DAX made a gain of 12% in the randomly chosen year, you would receive an interest rate of 15% on your investment (with 115 Euros earned for each 100 Euros of your investment).
- If, in the randomly chosen year, the DAX made a loss of 12%, you would receive a negative interest rate, which however would be lower due to the bonus of 3%, i.e. a loss of 9%. In this case you would earn 91 Euros for each 100 euros invested.

In this example your complete payment would be made up of 14.000 Euros (the fixed payment) plus the result of the investment in the Bund plus the result of the investment in the fund.

Please note that the bonus on the Bund later may differ from the bonus on the fund. They are merely equally high in this example.

(The example is of course arbitrary and does not contain information on the actual development of the DAX.)

Screen 3

You will receive the total payment of one of the 8 rounds on a smaller scale. At the end of the experiment the computer will choose one of the 8 rounds at random. Every round has the same probability of being chosen. This round will be paid out in cash.

Moreover, the computer randomly chooses a year from 1951 to 2010, also with equal probability. The gain or loss on the DAX in this year will be used to determine your payment.

For every 5000 Euro that you obtain in this round you will receive 1 Euro in cash.

To sum up: The Bund yields an interest of 2% plus the corresponding bonus while the fund can yield every DAX-profit or DAX-loss of the years 1951 to 2010 plus the corresponding bonus.

Investment Screen

Round 1

You have 50,000 Euros which you have to distribute over the Bund and the fund. In addition, you will receive 17,550 Euros independent of your choice.

For the Bund, the bonus is 2.80 percentage points.

For the fund, the bonus is 5.90 percentage points.

Bonus on the Bund	Bonus on the fund	Endowment	Additional payment
2.80 percentage points	5.90 percentage points	50,000 Euro	17,550 Euro

How much of the 50,000 EUR do you want to invest in the Bund and how much do you want to invest in the fund?

Please make sure that both amounts are integers and sum up to exactly 50,000 EUR.

Bund ....Euro
fund ....Euro

Screen 4

As you know, the return on the fund depends on the development of the DAX in the years from 1951 to 2010. In the following, we want to ask you what you think the DAX's gains and losses were during this period of time.

Therefore we will group the possible gains and losses of the fund in seven ranges on the next screen.:

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

On the next screen there are 20 boxes above each of these seven ranges. Please show us for the seven ranges how often you expect the DAX to have been in each range by clicking the mentioned boxes.

Please mark exactly twenty boxes. One box stands for a frequency of 1 in 20, i.e. for 5 percent.

By marking the boxes you will show us how likely you believe it is that your fund will produce a gain or loss in the given range.

- If, for instance, you don't mark any of the boxes in a particular range, this will mean that you are sure that the gain or loss will never lie in this range.
- If you mark one or two boxes in a particular range, you believe a loss or gain in this range to be possible, but not very likely.
- More boxes — up to 20 in one range — imply correspondingly higher probabilities.

Screen 5

Please mark the 20 boxes according to your assessment of the development of the DAX in the years from 1951 to 2010.

Mark the boxes, avoiding gaps from BOTTOM to TOP!

Loss of 60% to 90%	Loss of 30% to 90%	Loss of 0% to 30%	Gain of 0% to 30%	Gain of 30% to 60%	Gain of 60% to 90%	Gain of 90% to 120%

Overview of Payoffs

That's it. Thank you for participating!

The computer has determined by random draw that you will receive your investment of round 6. In this round you invested 25,000 EUR in the Bund and 25,000 EUR in the DAX. In this round the bonus on the interest rate of the Bund was 3.00 percentage points, and the bonus on the DAX was 6.05 percentage points.

Moreover, the computer has randomly chosen the year 1992. In this year the DAX made a loss of 0.66%.

As a result, we will pay you 14 EUR based on your investment, which are computed as follows:

Asset	Invested amount	Gain / Loss	Bonus	Overall Gain/ Loss	Payoff
Bund	25,000 EUR	2 %	3.00 %	5.00 %	26,250 EUR
Fund	25,000 EUR	-0.66 %	6.05 %	5.39 %	26,348 EUR
				Additional Payment	15,800 EUR
				Total	68,398 EUR

Total / 5000	13.68 EUR
(Rounded up to the nearest Euro)	14 EUR

Please remain seated for a little while. We will start the payment as soon as the vast majority of participants has completed the experiment.

S.2.5 Decision Screen in Complexity Experiment

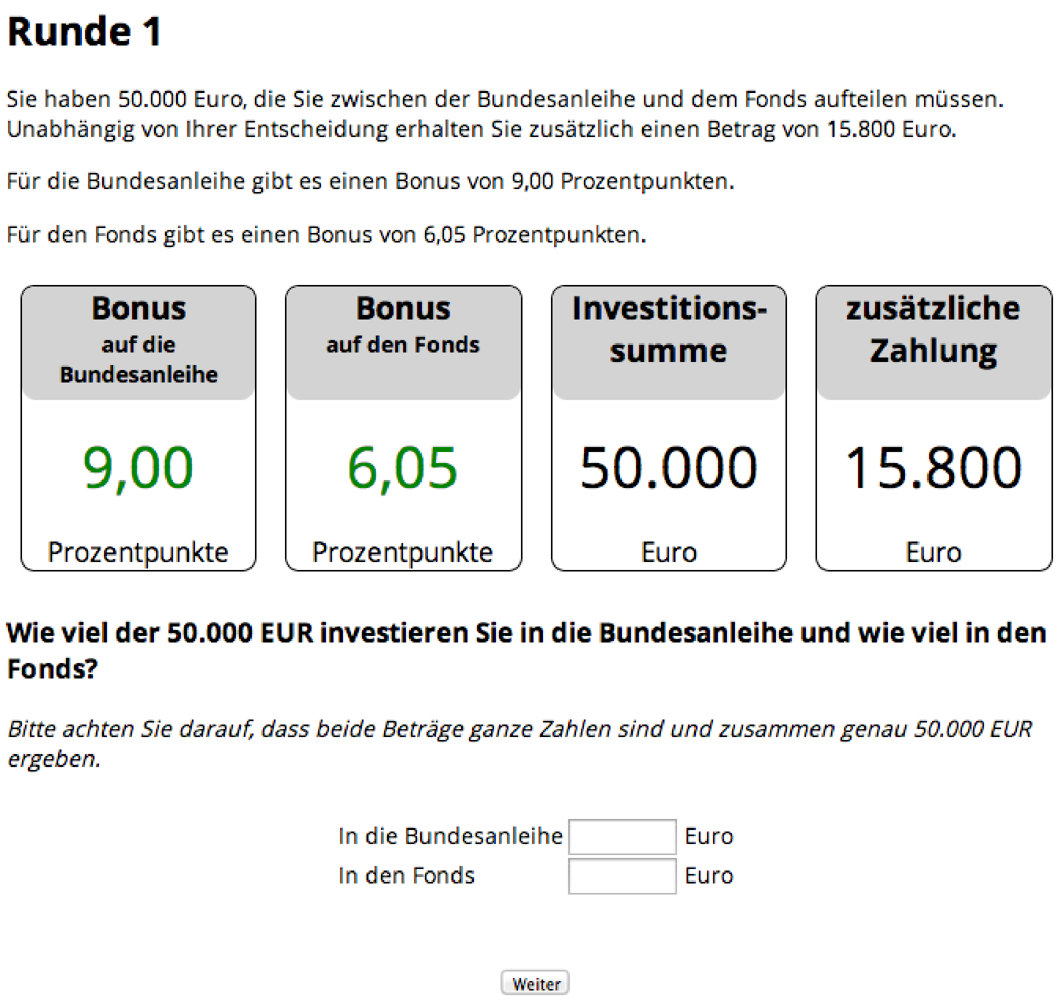


Figure S.5: Decision Screen

S.3 Measuring Ambiguity Aversion: Experimental Tests of Subjective Expected Utility

S.3.1 Experimental Instructions

Hello and Welcome.

ELIGIBILITY FOR THIS STUDY:

This study is totally anonymous. You must be willing to receive your payment for this study by cash at the end of the experiment today.

EARNING MONEY:

To begin, you will be given a $5 minimum payment. Whatever you earn from the study today will be added to this minimum payment.

In this study, you will make a series of choices between two jars, each containing 20 marbles. The first jar will always be called JAR A. JAR A will contain green marbles worth either $10 or $30 and red marbles worth either $10 or $30.

The second jar will always be called JAR B. JAR B will contain yellow marbles worth $0 and black marbles worth $30. Yellow marbles will always be worth $0 and black marbles will always be worth $30.

For each decision, your task is to decide whether you prefer to draw a marble from JAR A or draw a marble from JAR B.

Once all of your decisions have been made, we will randomly select one decision as the *decision-that-counts*. If you preferred JAR A, then we will draw a ball from JAR A to determine your earnings while if you preferred JAR B we will draw a ball from JAR B instead.

JAR A, JAR B or both will involve risk. That is, there may be some green marbles and some red marbles in JAR A. And, there may be some black marbles and some yellow marbles in JAR B.

For example, JAR A might contain 4 green marbles worth $30 and 16 red marbles worth $10. JAR B might contain 12 black marbles worth $30 and 8 yellow marbles worth $0. Imagine for a second which jar you would rather draw a marble from. You have been provided with a calculator to help you in making your decisions.

Additionally, some of the contents of JAR A may be unknown to you. That is, you will know that there is a mix of red and green marbles, but you will not know the exact proportions. For example, JAR A might contain 4 green marbles worth $30, 6 red marbles worth $10 and 10 unknown marbles, which could be red marbles, green marbles, or some mixture of the two. These unknown marbles will be represented as white marbles, such that the contents of JAR A would be described as containing 4 green marbles, 6 red marbles and 10 unknown, white marbles.

JAR B will never be unknown. You will always know the contents of JAR B.

Once we know which is the decision-that-counts, and whether you prefer JAR A or JAR B, we will then determine the value of your payments.

Let's take the example of the partially unknown JAR A just described containing 4 green marbles worth \$30, 6 red marbles worth \$10, and 10 unknown white marbles either \$30 green marbles or \$10 red marbles. Suppose this is compared to the JAR B previously described, containing 12 black marbles worth \$30 and 8 yellow marbles worth \$0.

If this was chosen as the decision-that-counts, we would prepare JAR A and JAR B.

You will watch as we place 12 green marbles and 8 yellow marbles in JAR B on the table at the front of the room.

If you preferred JAR B, we will draw a marble from JAR B. If the drawn marble is black you will receive \$30 (+5 minimum payment) = \$35. If the drawn marble is yellow you will receive \$0 (+5 minimum payment) = \$5.

You will also watch as we place 4 green marbles, 6 red marbles, and 10 white marbles in JAR A on the table at the front of the room.

If you preferred JAR A, we will draw a marble from JAR A. If the drawn marble is green you will receive \$30 (+5 minimum payment) = \$35. If the drawn marble is red you will receive \$10 (+5 minimum payment) = \$15.

If the marble drawn from JAR A is one of the white unknown marbles, however, the following procedure will determine whether the marble is green or red. There is a third jar on the table at the front of the room right now, marked JAR X. JAR X contains 20 marbles. These marbles are some combination of red and green and have already been determined. There may be anywhere from 0 red marbles and 20 green marbles to 20 red marbles and 0 green marbles in JAR X, or any combination of 20 red and green marbles.

If a white marble is drawn from JAR A, we will draw a marble from JAR X. If the marble drawn from JAR X is green you will receive the payment for a green marble (here it is \$30 (+5 minimum payment) = \$35). If the marble drawn from JAR X is red you will receive the payment for a red marble (here it is \$10 (+5 minimum payment) = \$15). A number of questions will refer to the unknown white marbles and so refer to JAR X. You should think carefully about the possible contents of JAR X.

THINGS TO REMEMBER:

1. You will receive a \$5 minimum payment just for participating today.
2. You will complete 20 decision tasks
3. In each decision you will be asked whether you prefer to draw a marble from JAR A or JAR B. Your task is decide whether you prefer JAR A or JAR B.
4. JAR B will contain Yellow marbles worth \$0 and Black marbles worth \$30.
5. JAR A will contain Red and Green marbles worth either \$30 or \$10. Some of the contents of JAR A may be unknown to you. These unknown marbles will be described as white marbles.
6. White marbles will entail a draw from JAR X. The contents of JAR X are unknown to you. They could be any combination of red and green marbles.
7. At the end of the study one decision will be chosen as the decision-that-counts.
8. Because each decision is equally likely, you should make each decision as if it is the one that determines your payments.
9. Once we know the decision that counts, we will construct JAR A and JAR B. If you preferred JAR B in the decision-that-counts, we would draw a marble from JAR B and pay you accordingly. If you preferred JAR A in the decision-that-counts, we would draw a marble from JAR A and pay you accordingly. If the marble drawn from JAR A is a white marble, a marble will be drawn from JAR X to determine your payments.

In a moment, we will proceed to the tasks. You have been given a subject ID number. Please write it on the top of your task packet.

2 First Task Block: Green Pays

DECISION BLOCK 1

Participant Number:

TASKS 1-8

On the following pages you will complete 8 tasks. In each task you are asked to make a series of decisions between two jars: JAR A and JAR B

JAR A will contain Red Marbles, Green Marbles and Unknown White Marbles which could be either Red, Green or some mixture between the two. **Red Marbles in the following 8 tasks will be worth $10. Green Marbles in the following 8 tasks will be worth $30.** JAR B will contain Yellow Marbles worth $0 and Black Marbles worth $30.

In each task, JAR A will be fixed, while JAR B will vary. For example, in Task 1 JAR A will contain 0 $10 Red Marbles, 15 $30 Green Marbles and 5 unknown White Marbles, which could be red, green or some mixture of the two. This will remain the same for all decisions in the task.

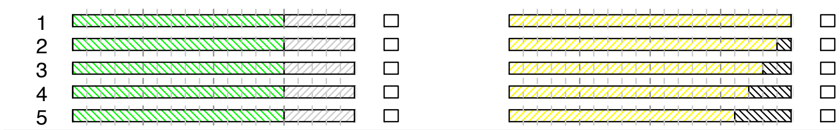
JAR B will vary across decisions. Initially JAR B will contain 20 $0 Yellow Marbles and 0 $30 Black Marbles. As you proceed, the contents of JAR B will change. The number of Black $30 marbles will increase while the number of Yellow $0 marbles will decrease. In each subsequent row, one Yellow $0 marble will be removed and one Black $30 marble will be added.

For each row, your task is to decide whether you prefer to draw a marble from JAR A or JAR B. Indicate your preference by checking the corresponding box.

The first several decisions from Task 1 are reproduced as an example.

EXAMPLE

JAR A	**JAR B**
0 Red Marbles worth $10 each and 15 Green Marbles worth $30 each and 5 Unknown White Marbles	Yellow Marbles worth $0 each and Black Marbles worth $30 each.

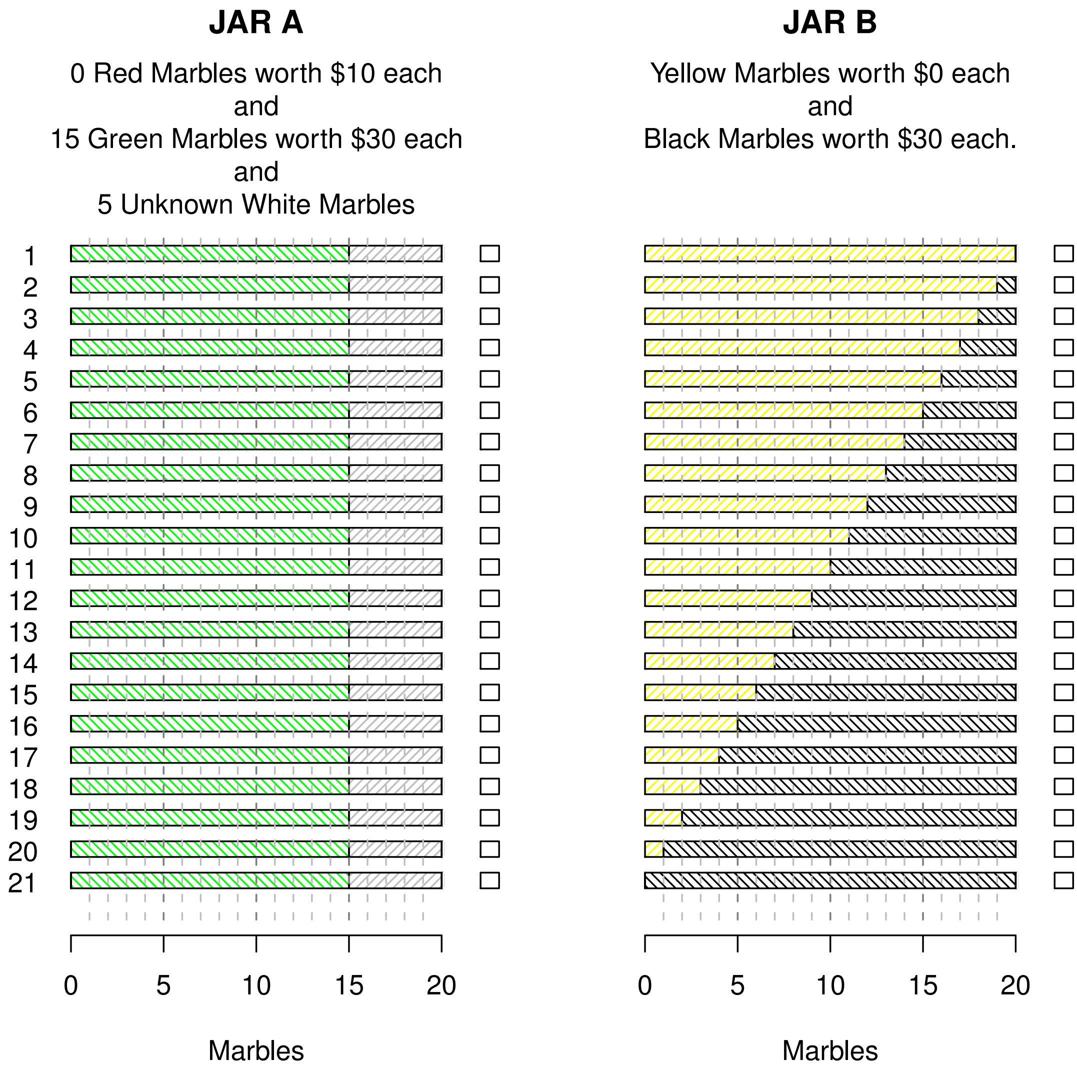


In the first decision you are asked to decide between a draw from JAR A which contains 0 $10 Red Marbles, 15 $30 Green Marbles and 5 unknown White Marbles, and a draw from JAR B which contains 20 $0 Yellow Marbles and 0 $30 Black Marbles. If you prefer a draw from JAR A, check the left-hand box under JAR A. If your prefer a draw from JAR B, check the right-hand box under JAR B.

Each decision could be the **decision-that-counts**. So, it is in your interest to treat each decision as if it could be the one that determines your payments.

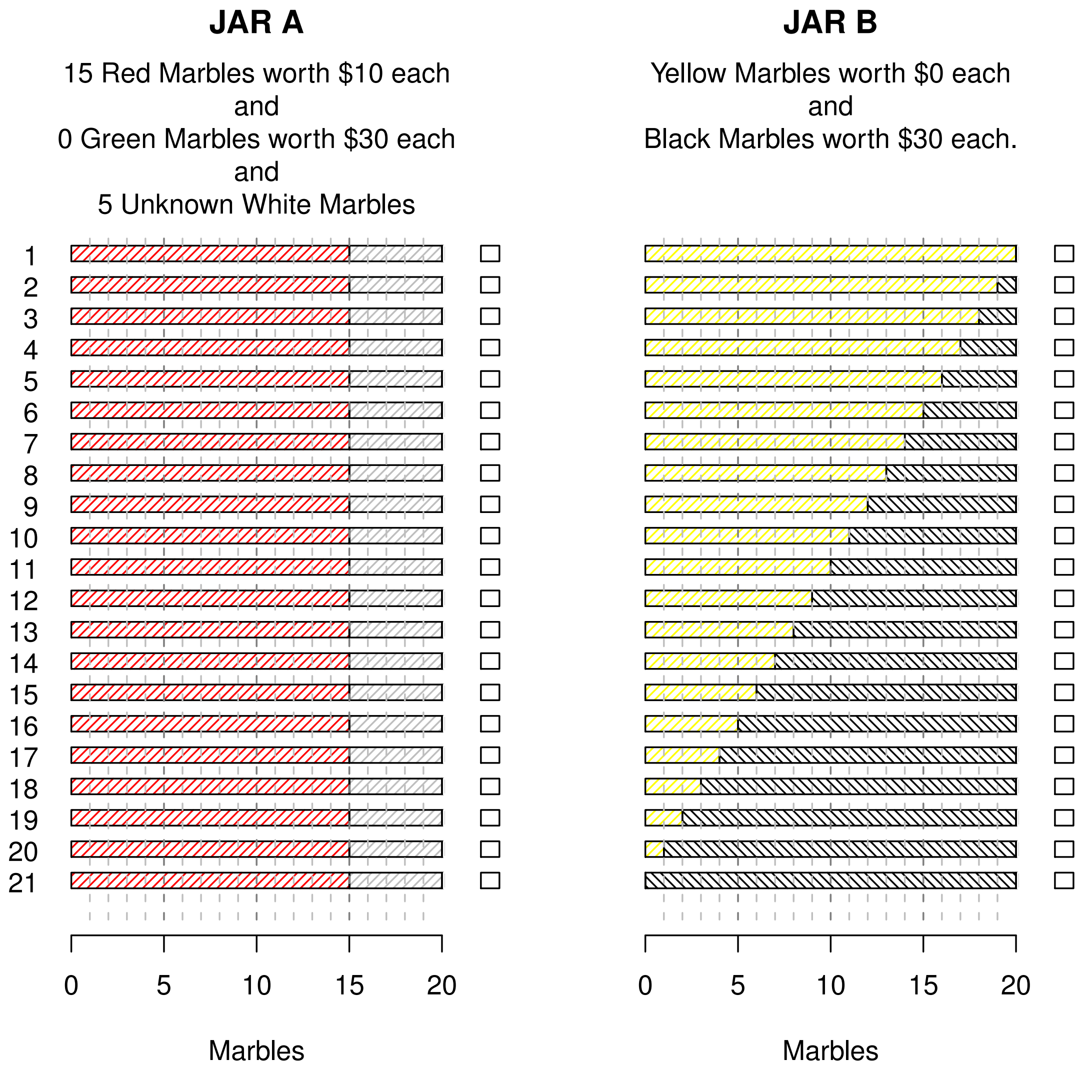
TASK 1

On this page you will make a series of decisions between two jars. JAR A will contain 0 Red marbles worth $10 each, 15 Green marbles worth $30 each, and 5 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



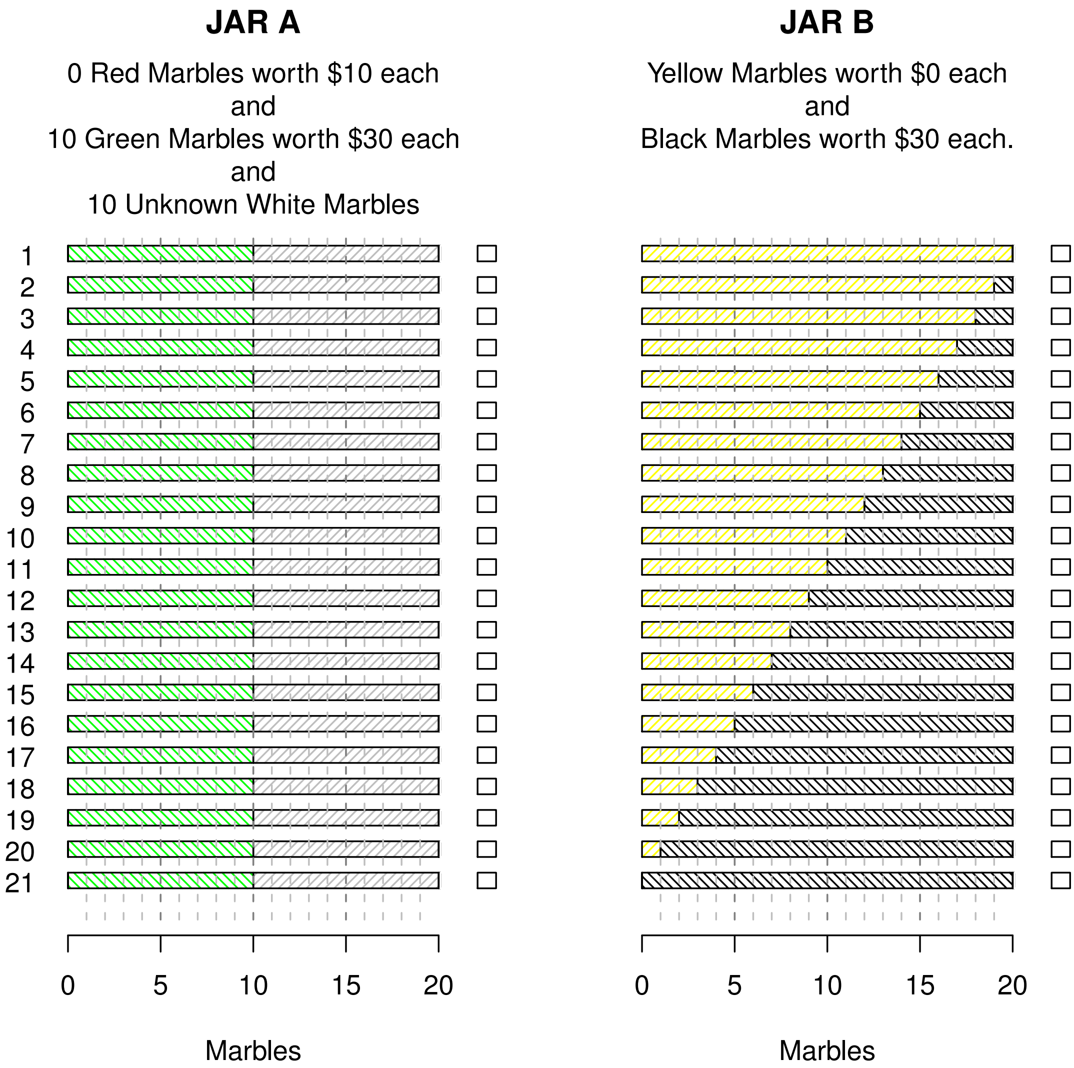
TASK 2

On this page you will make a series of decisions between two jars. JAR A will contain 15 Red marbles worth $10 each, 0 Green marbles worth $30 each, and 5 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



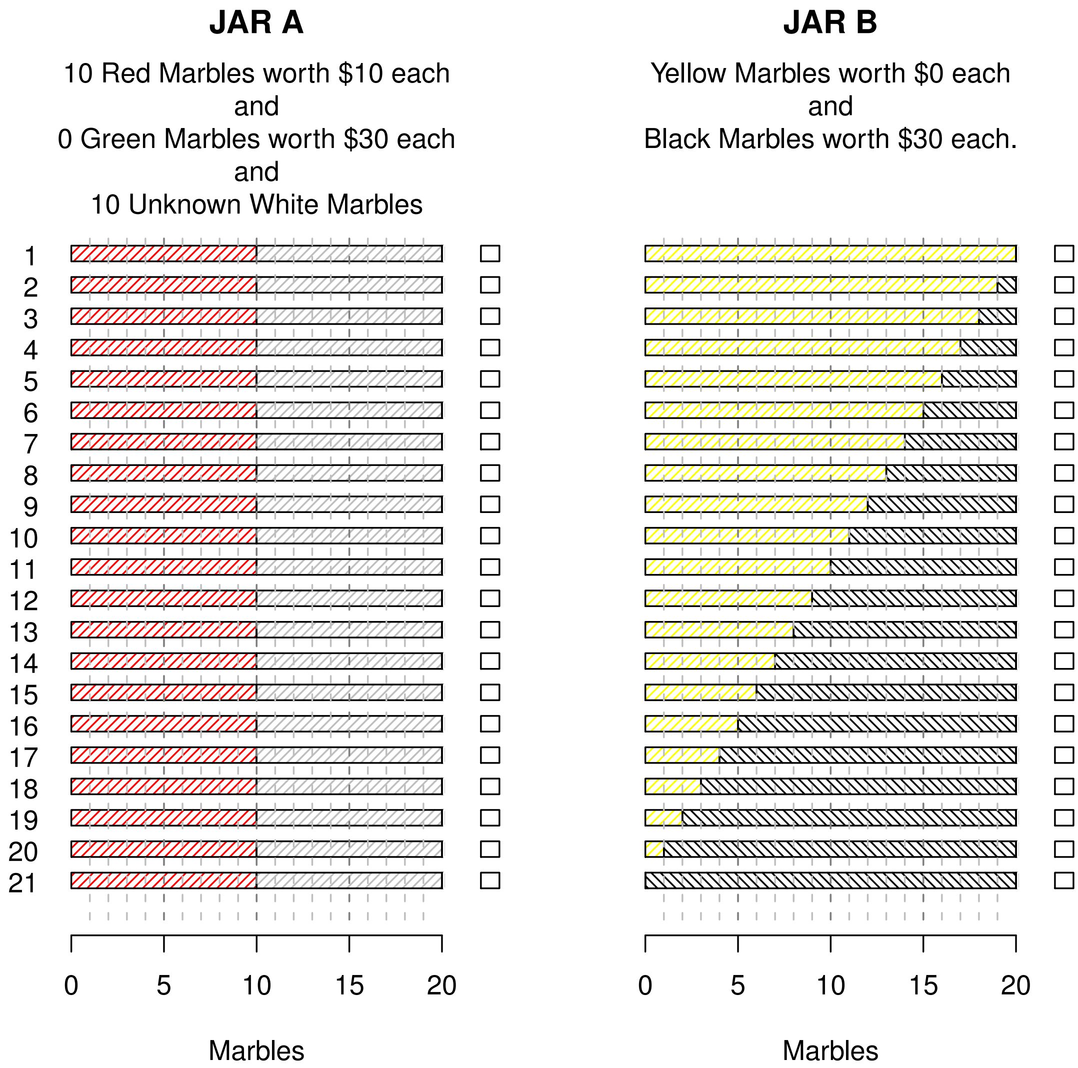
TASK 3

On this page you will make a series of decisions between two jars. JAR A will contain 0 Red marbles worth $10 each, 10 Green marbles worth $30 each, and 10 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



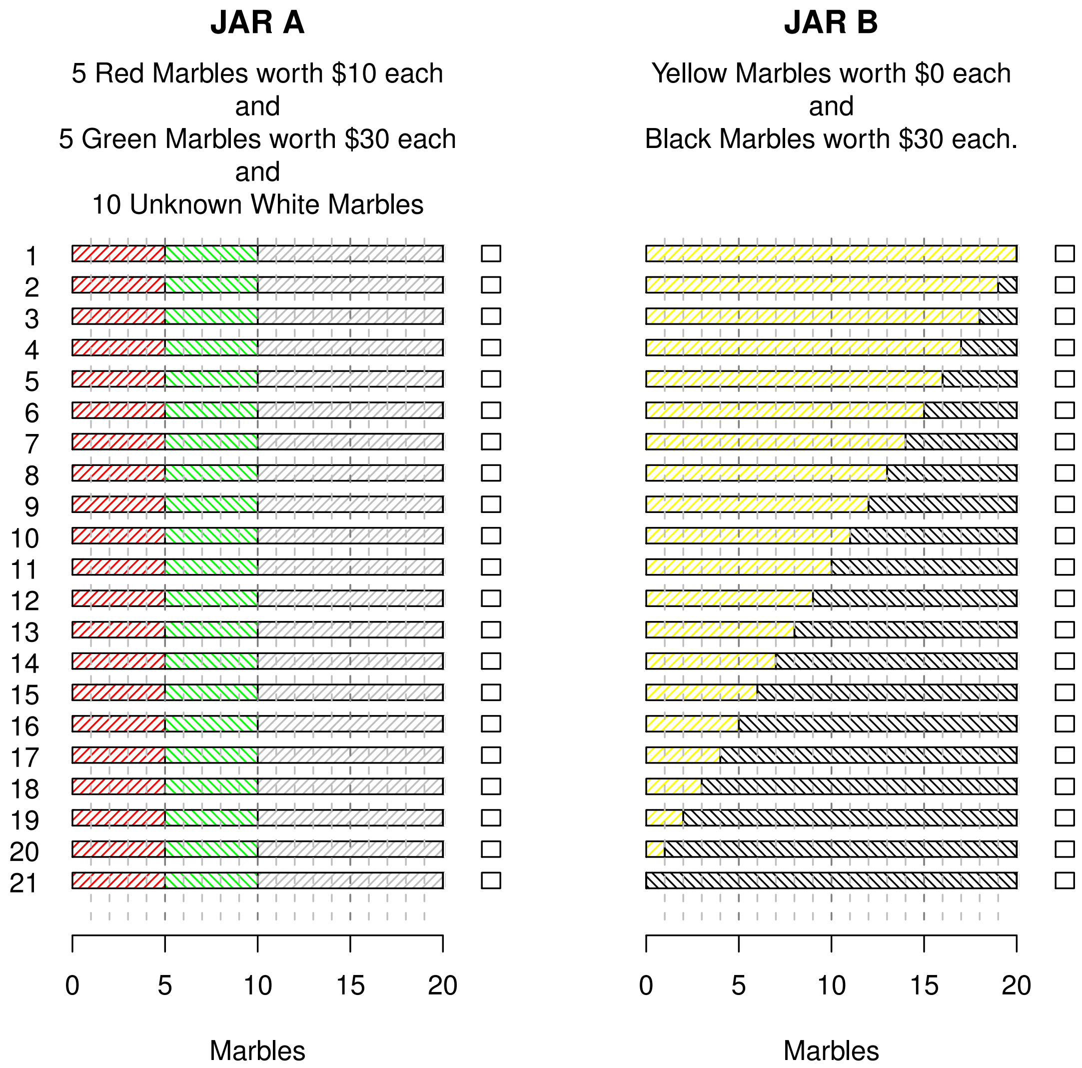
TASK 4

On this page you will make a series of decisions between two jars. JAR A will contain 10 Red marbles worth \$10 each, 0 Green marbles worth \$30 each, and 10 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



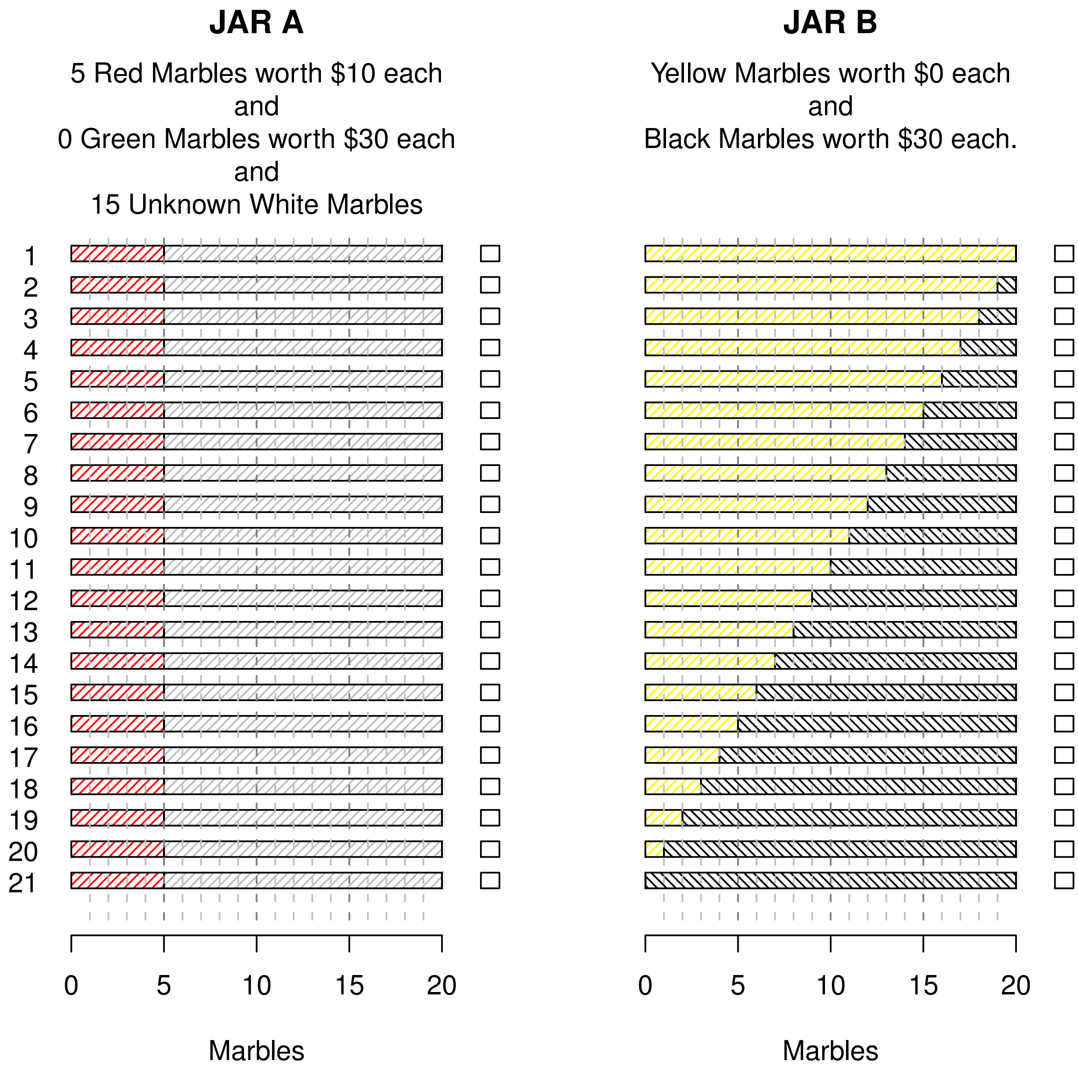
TASK 5

On this page you will make a series of decisions between two jars. JAR A will contain 5 Red marbles worth $10 each, 5 Green marbles worth $30 each, and 10 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



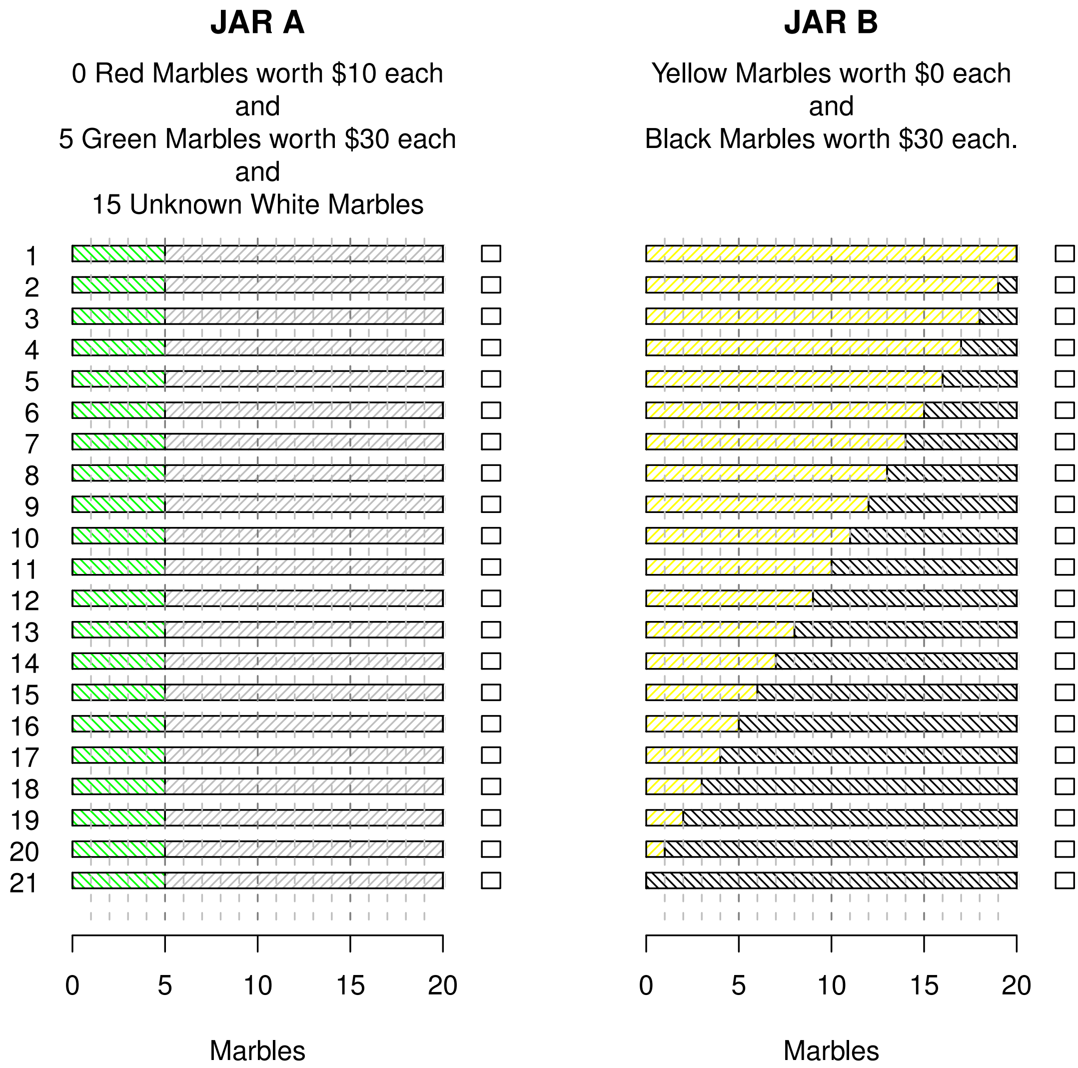
TASK 6

On this page you will make a series of decisions between two jars. JAR A will contain 5 Red marbles worth $10 each, 0 Green marbles worth $30 each, and 15 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



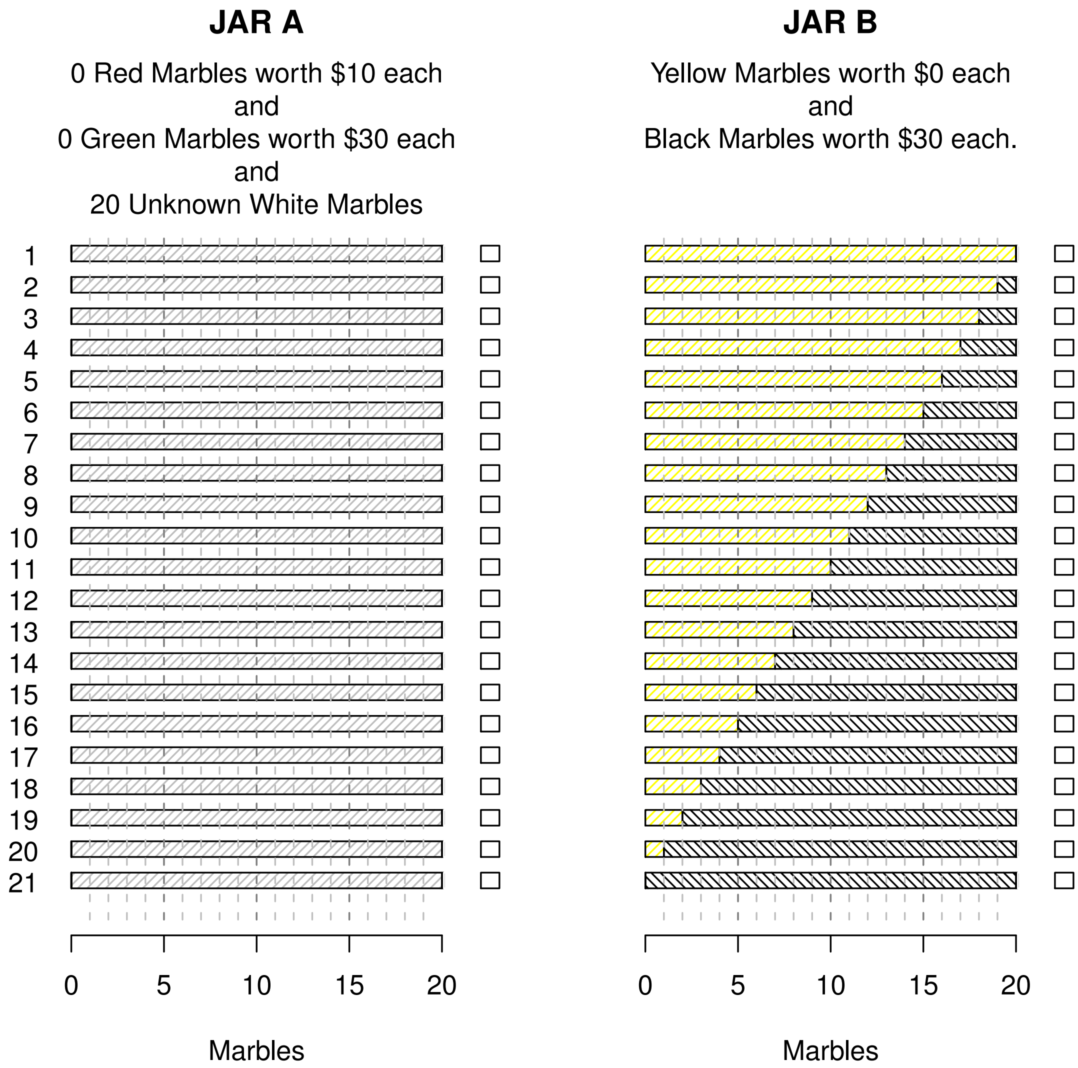
TASK 7

On this page you will make a series of decisions between two jars. JAR A will contain 0 Red marbles worth $10 each, 5 Green marbles worth $30 each, and 15 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



TASK 8

On this page you will make a series of decisions between two jars. JAR A will contain 0 Red marbles worth $10 each, 0 Green marbles worth $30 each, and 20 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



3 Second Task Block: No Ambiguity

DECISION BLOCK 2

Participant Number:

TASKS 9-12

On the following pages you will complete 4 tasks. In each task you are asked to make a series of decisions between two jars: JAR A and JAR B

JAR A will contain Red Marbles, Green Marbles and Unknown White Marbles which could be either Red, Green or some mixture between the two. **Red Marbles in the following 4 tasks will be worth $10. Green Marbles in the following 4 tasks will be worth $30.** JAR B will contain Yellow Marbles worth $0 and Black Marbles worth $30.

In each task, JAR A will be fixed, while JAR B will vary. For example, in Task 9 JAR A will contain 5 $10 Red Marbles, 15 $30 Green Marbles and 0 unknown White Marbles, which could be red, green or some mixture of the two. This will remain the same for all decisions in the task.

JAR B will vary across decisions. Initially JAR B will contain 20 $0 Yellow Marbles and 0 $30 Black Marbles. As you proceed, the contents of JAR B will change. The number of Black $30 marbles will increase while the number of Yellow $0 marbles will decrease. In each subsequent row, one Yellow $0 marble will be removed and one Black $30 marble will be added.

For each row, your task is to decide whether you prefer to draw a marble from JAR A or JAR B. Indicate your preference by checking the corresponding box.

The first several decisions from Task 9 are reproduced as an example.

EXAMPLE

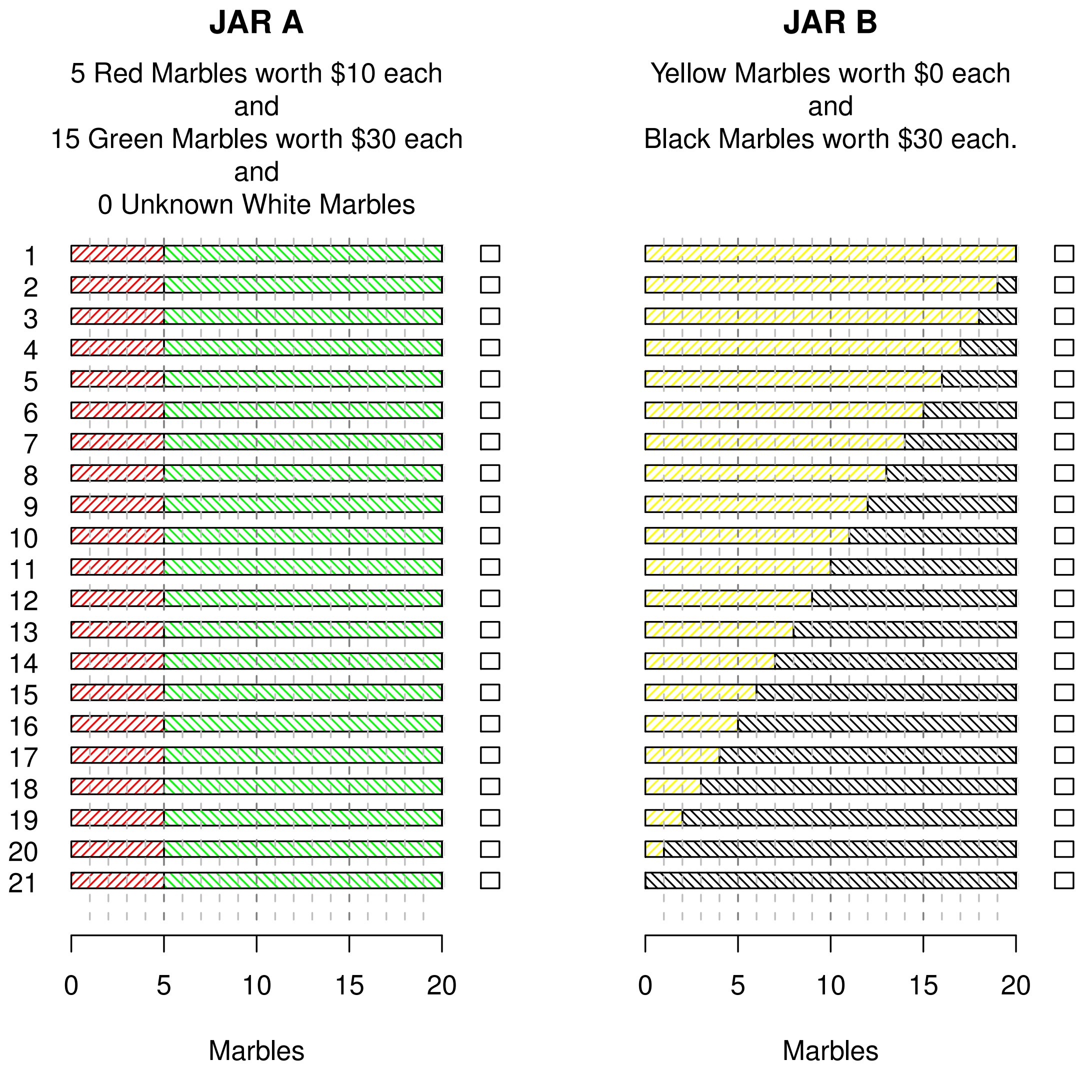
JAR A	**JAR B**
5 Red Marbles worth $10 each and 15 Green Marbles worth $30 each and 0 Unknown White Marbles	Yellow Marbles worth $0 each and Black Marbles worth $30 each.

In the first decision you are asked to decide between a draw from JAR A which contains 5 $10 Red Marbles, 15 $30 Green Marbles and 0 unknown White Marbles, and a draw from JAR B which contains 20 $0 Yellow Marbles and 0 $30 Black Marbles. If you prefer a draw from JAR A, check the left-hand box under JAR A. If your prefer a draw from JAR B, check the right-hand box under JAR B.

Each decision could be the **decision-that-counts**. So, it is in your interest to treat each decision as if it could be the one that determines your payments.

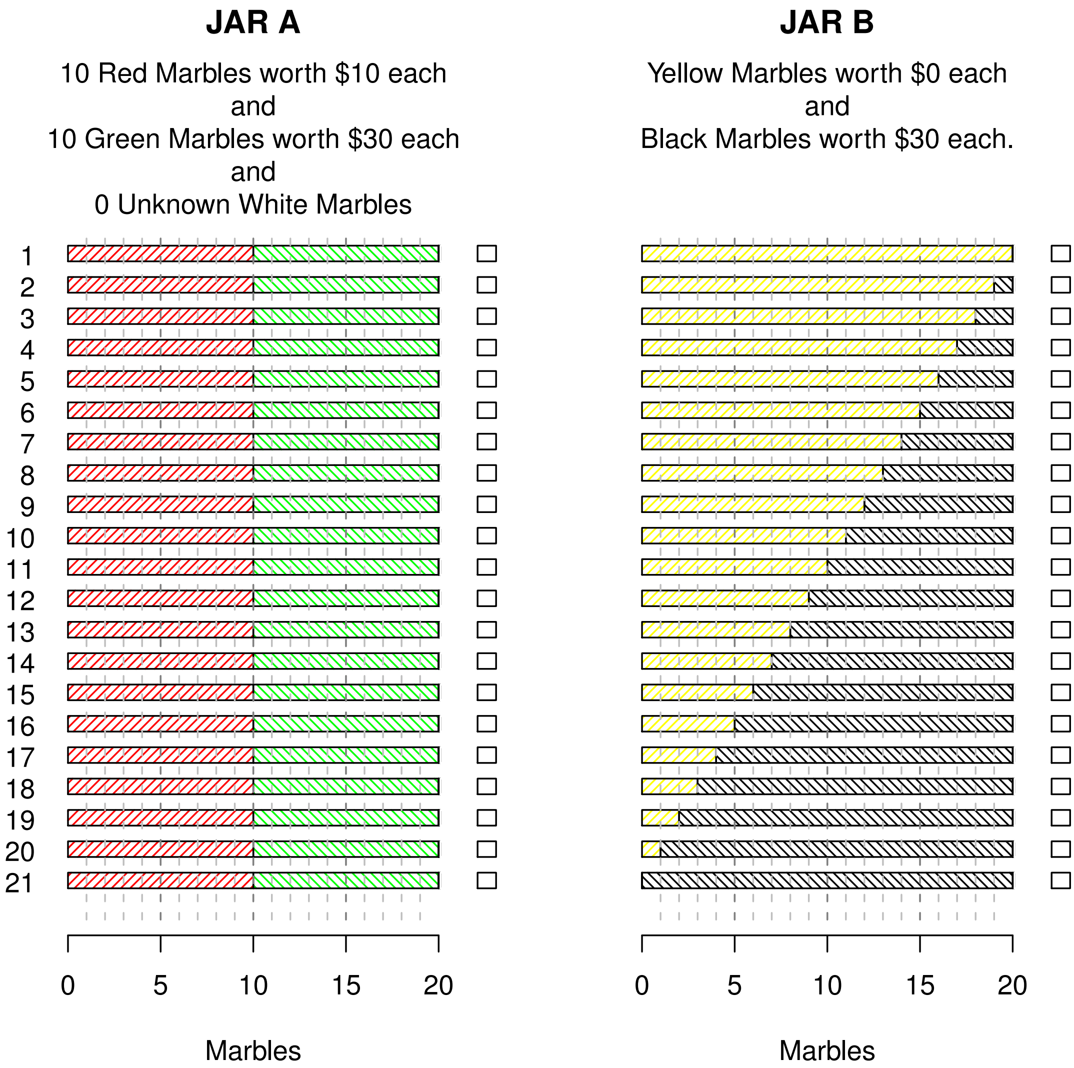
TASK 9

On this page you will make a series of decisions between two jars. JAR A will contain 5 Red marbles worth \$10 each, 15 Green marbles worth \$30 each, and 0 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



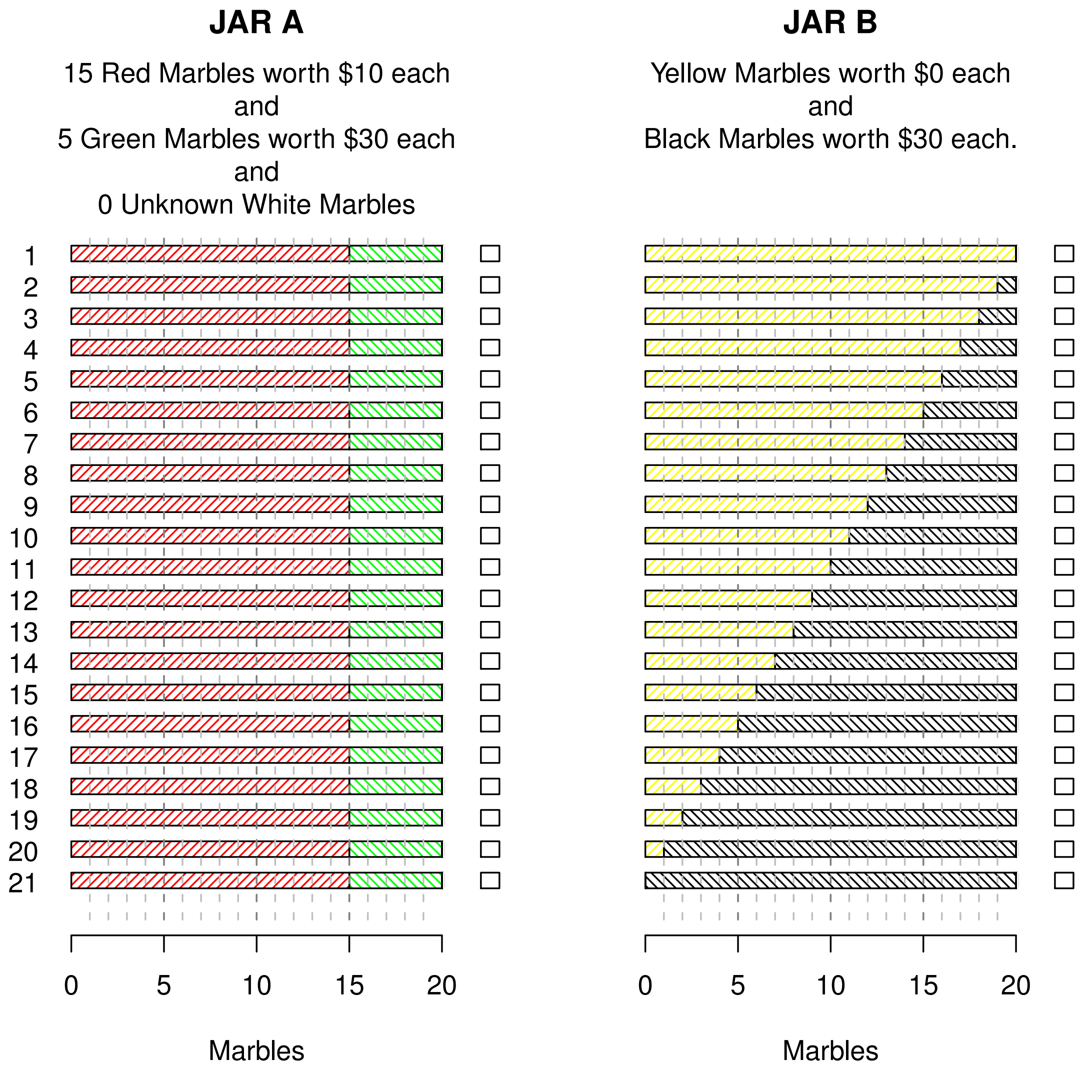
TASK 10

On this page you will make a series of decisions between two jars. JAR A will contain 10 Red marbles worth \$10 each, 10 Green marbles worth \$30 each, and 0 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



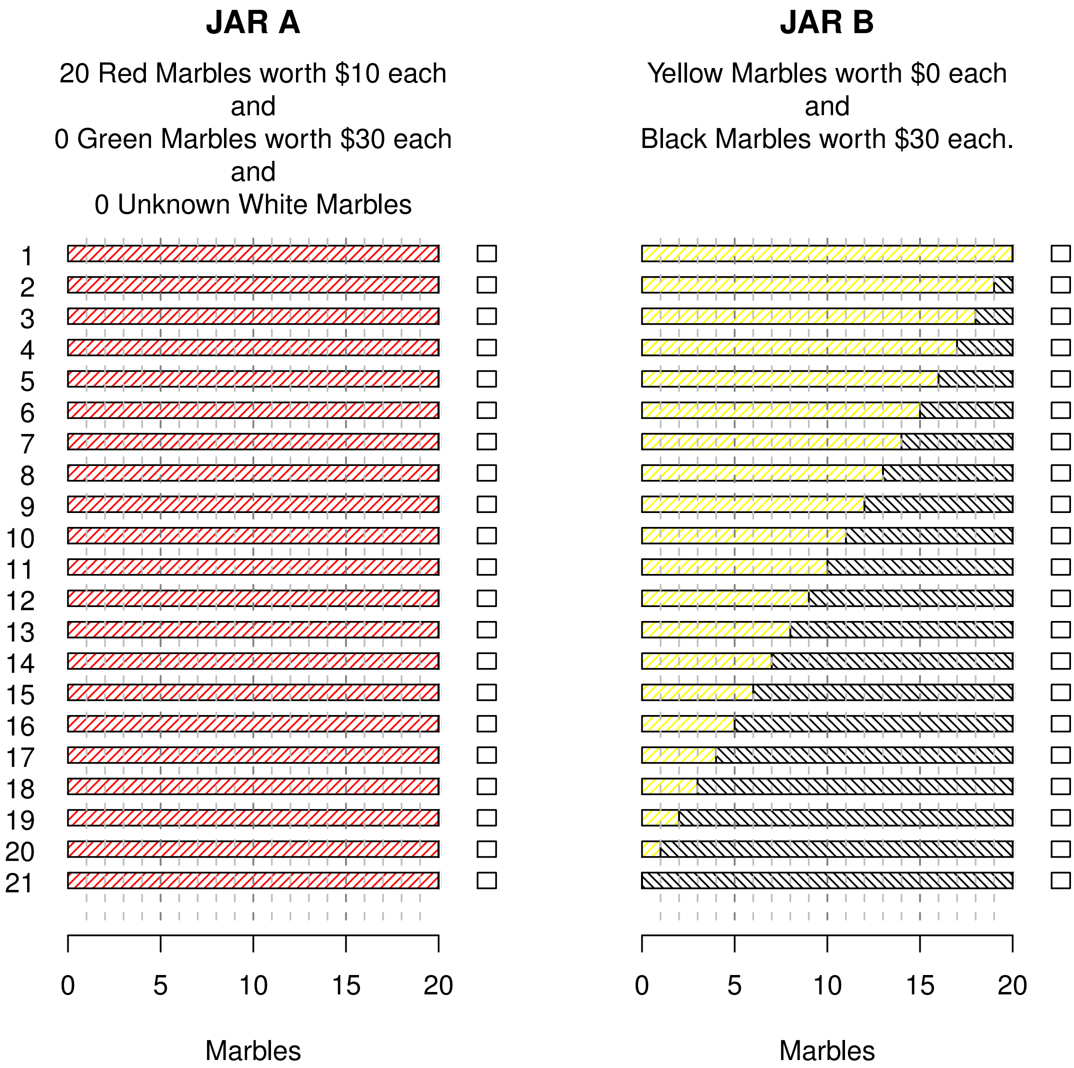
TASK 11

On this page you will make a series of decisions between two jars. JAR A will contain 15 Red marbles worth $10 each, 5 Green marbles worth $30 each, and 0 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



TASK 12

On this page you will make a series of decisions between two jars. JAR A will contain 20 Red marbles worth \$10 each, 0 Green marbles worth \$30 each, and 0 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



4 Third Task Block: Red Pays

DECISION BLOCK 3

Participant Number:

TASKS 13-20

On the following pages you will complete 4 tasks. In each task you are asked to make a series of decisions between two jars: JAR A and JAR B

JAR A will contain Red Marbles, Green Marbles and Unknown White Marbles which could be either Red, Green or some mixture between the two. **Green Marbles in the following 8 tasks will be worth $10. Red Marbles in the following 8 tasks will be worth $30.** JAR B will contain Yellow Marbles worth $0 and Black Marbles worth $30.

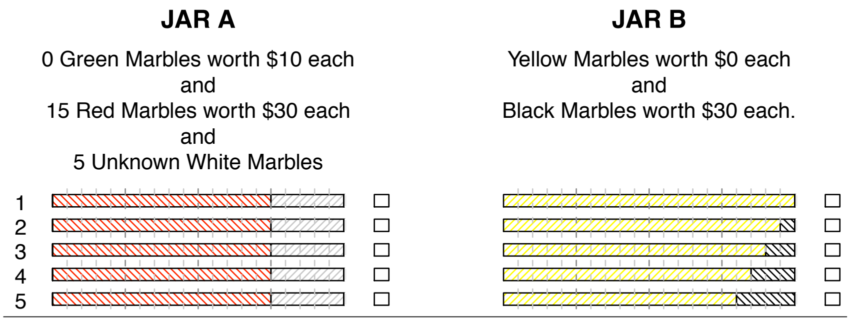
In each task, JAR A will be fixed, while JAR B will vary. For example, in Task 13 JAR A will contain 0 $10 Green Marbles, 15 $30 Red Marbles and 5 unknown White Marbles, which could be red, green or some mixture of the two. This will remain the same for all decisions in the task.

JAR B will vary across decisions. Initially JAR B will contain 20 $0 Yellow Marbles and 0 $30 Black Marbles. As you proceed, the contents of JAR B will change. The number of Black $30 marbles will increase while the number of Yellow $0 marbles will decrease. In each subsequent row, one Yellow $0 marble will be removed and one Black $30 marble will be added.

For each row, your task is to decide whether you prefer to draw a marble from JAR A or JAR B. Indicate your preference by checking the corresponding box.

The first several decisions from Task 13 are reproduced as an example.

EXAMPLE

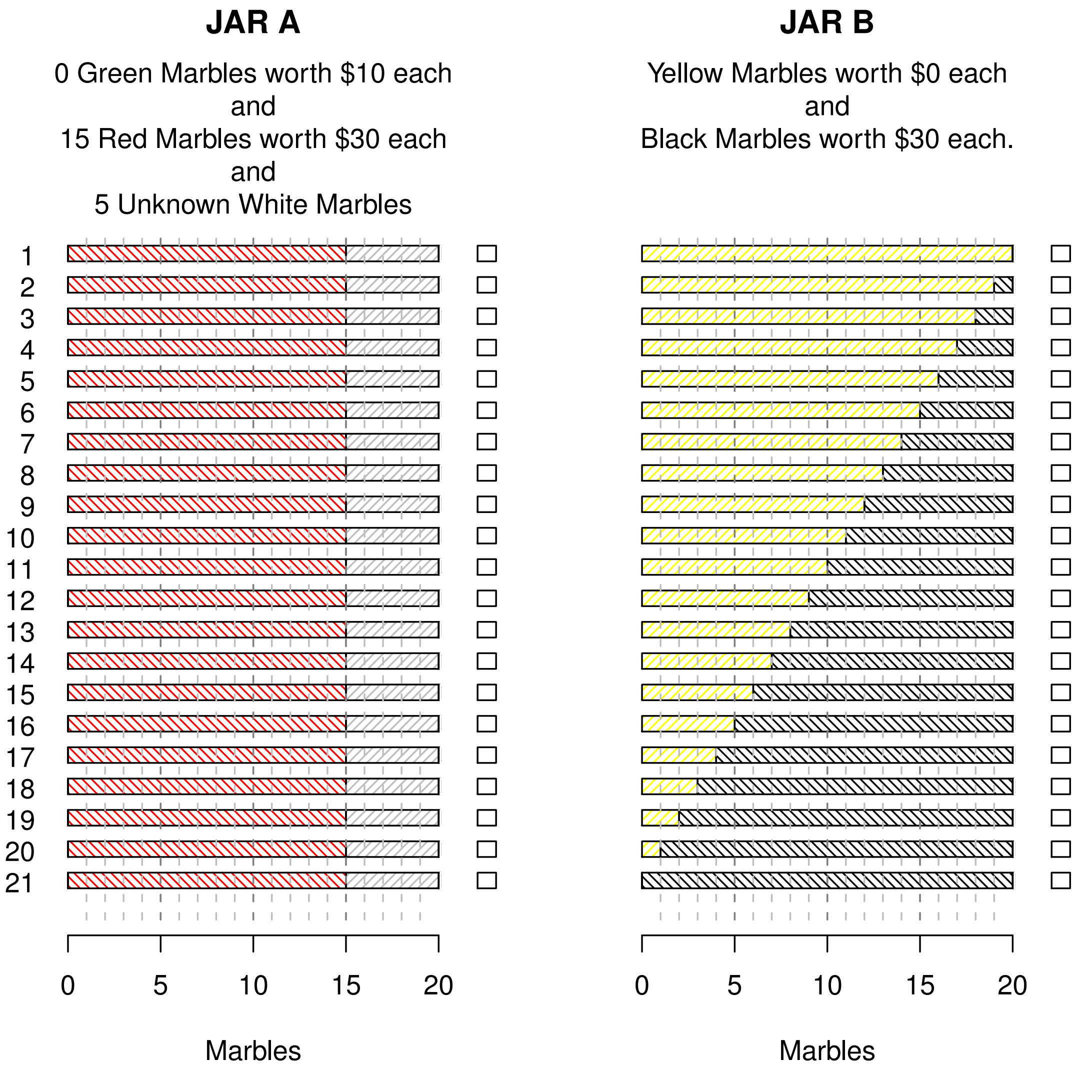


In the first decision you are asked to decide between a draw from JAR A which contains 0 $10 Green Marbles, 15 $30 Red Marbles and 5 unknown White Marbles, and a draw from JAR B which contains 20 $0 Yellow Marbles and 0 $30 Black Marbles. If you prefer a draw from JAR A, check the left-hand box under JAR A. If your prefer a draw from JAR B, check the right-hand box under JAR B.

Each decision could be the **decision-that-counts**. So, it is in your interest to treat each decision as if it could be the one that determines your payments.

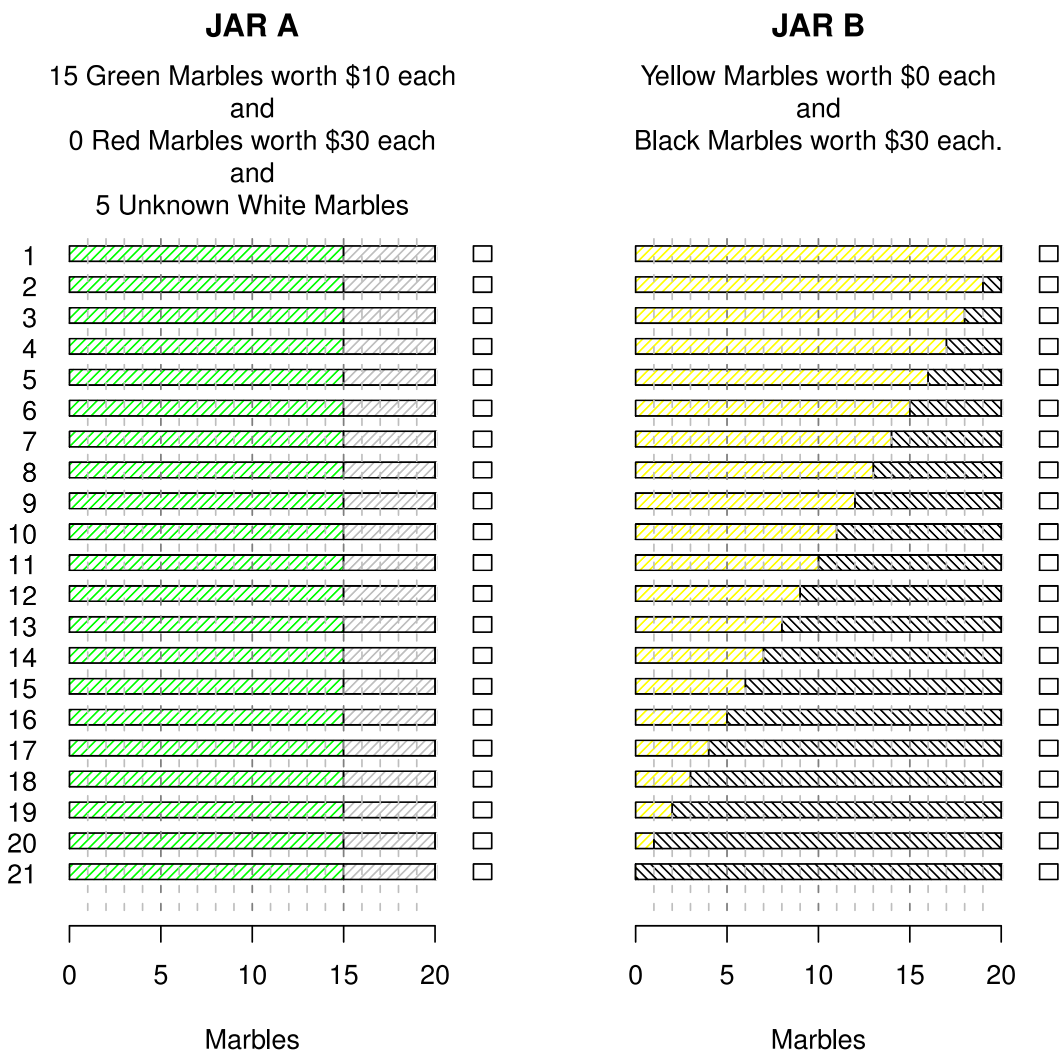
TASK 13

On this page you will make a series of decisions between two jars. JAR A will contain 0 Green marbles worth \$10 each, 15 Red marbles worth \$30 each, and 5 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



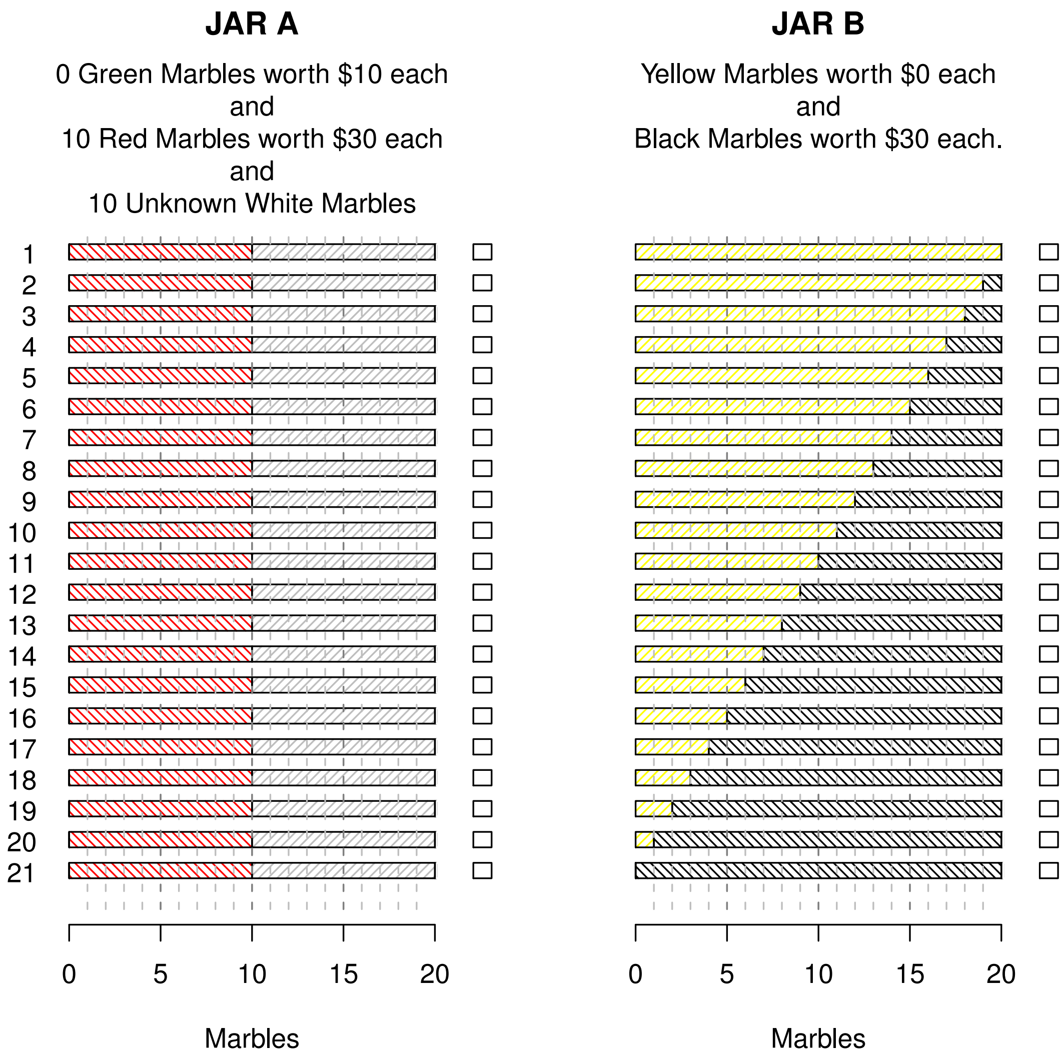
TASK 14

On this page you will make a series of decisions between two jars. JAR A will contain 15 Green marbles worth \$10 each, 0 Red marbles worth \$30 each, and 5 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



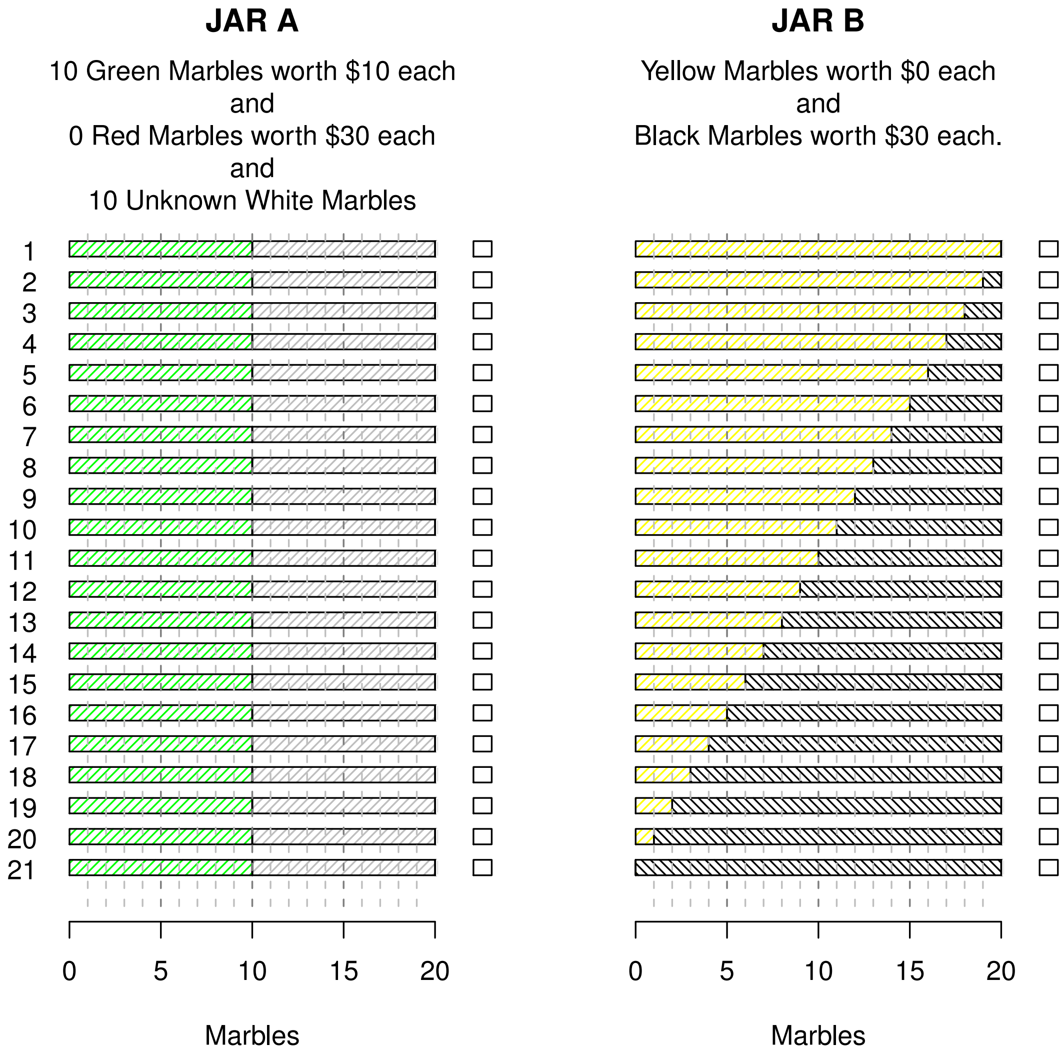
TASK 15

On this page you will make a series of decisions between two jars. JAR A will contain 0 Green marbles worth \$10 each, 10 Red marbles worth \$30 each, and 10 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



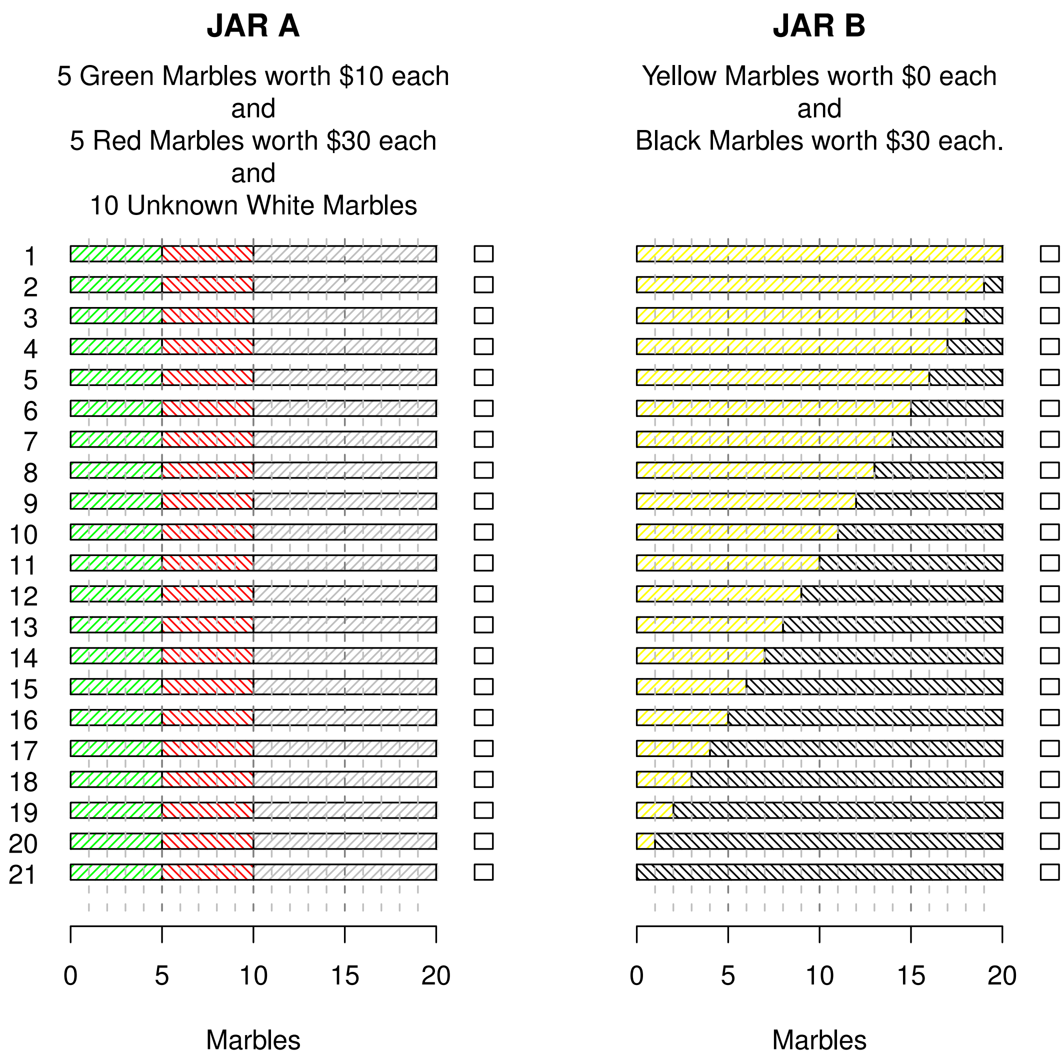
TASK 16

On this page you will make a series of decisions between two jars. JAR A will contain 10 Green marbles worth $10 each, 0 Red marbles worth $30 each, and 10 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



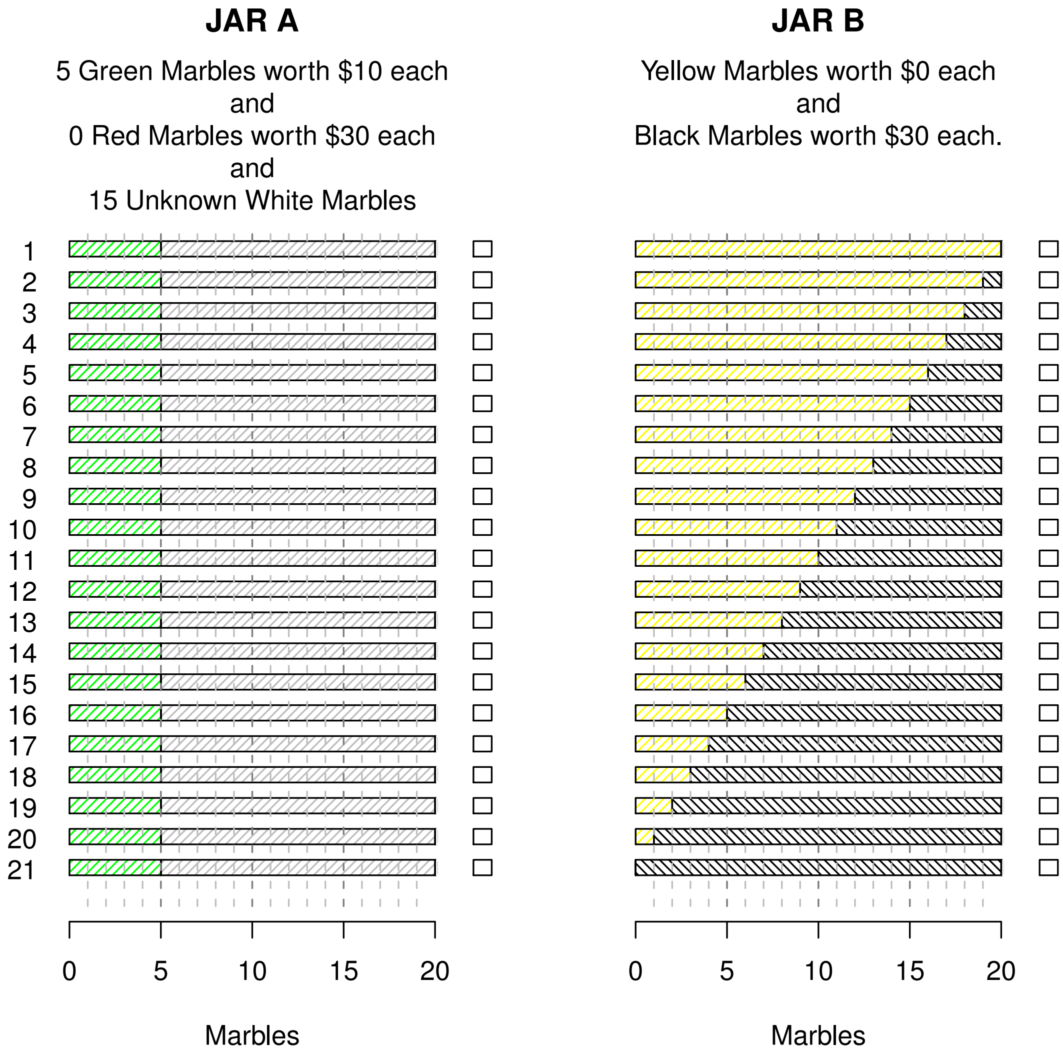
TASK 17

On this page you will make a series of decisions between two jars. JAR A will contain 5 Green marbles worth \$10 each, 5 Red marbles worth \$30 each, and 10 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



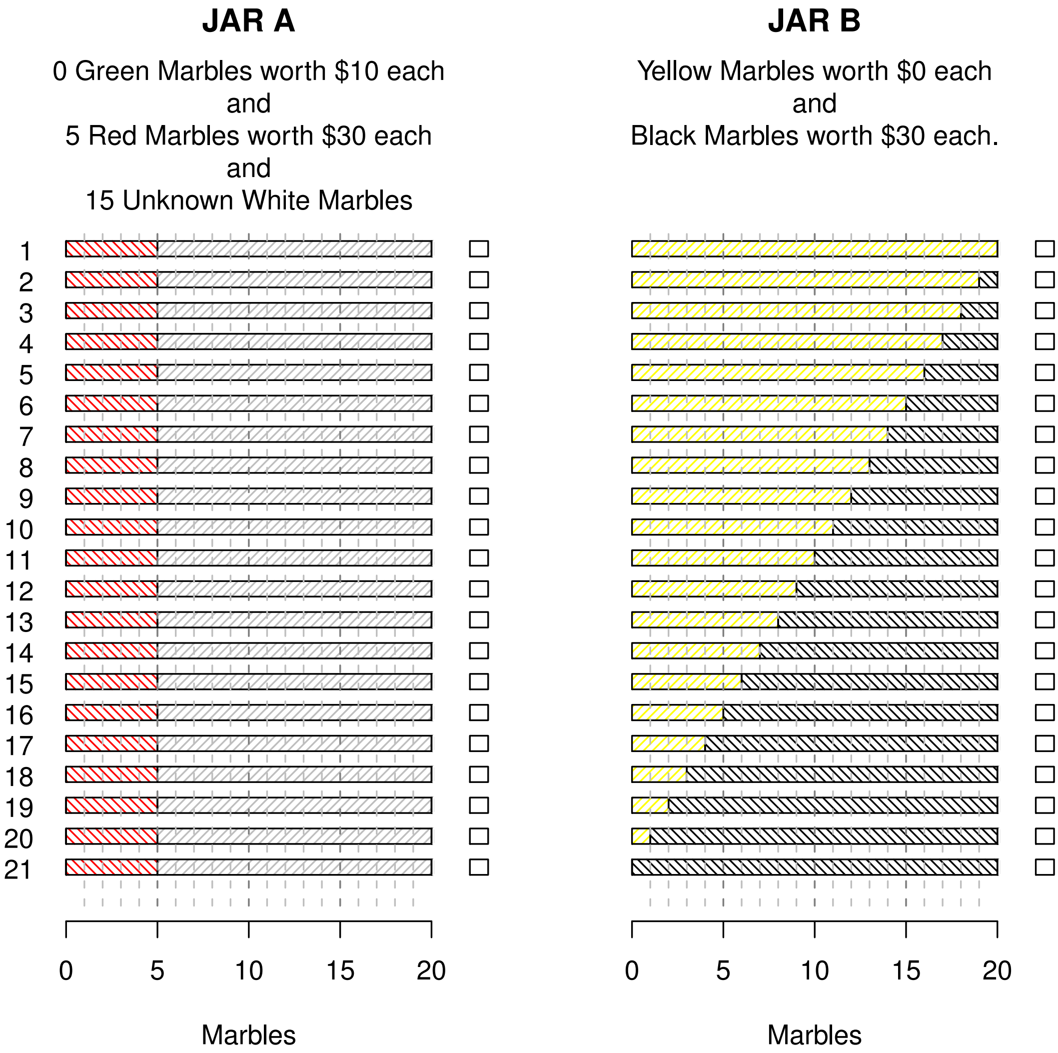
TASK 18

On this page you will make a series of decisions between two jars. JAR A will contain 5 Green marbles worth $10 each, 0 Red marbles worth $30 each, and 15 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth $0 each and Black marbles worth $30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow $0 Marbles and 0 Black $30 marbles. The contents of JAR B will change. When you move down one row, one Yellow $0 marble will be removed and replaced with one Black $30 marble.



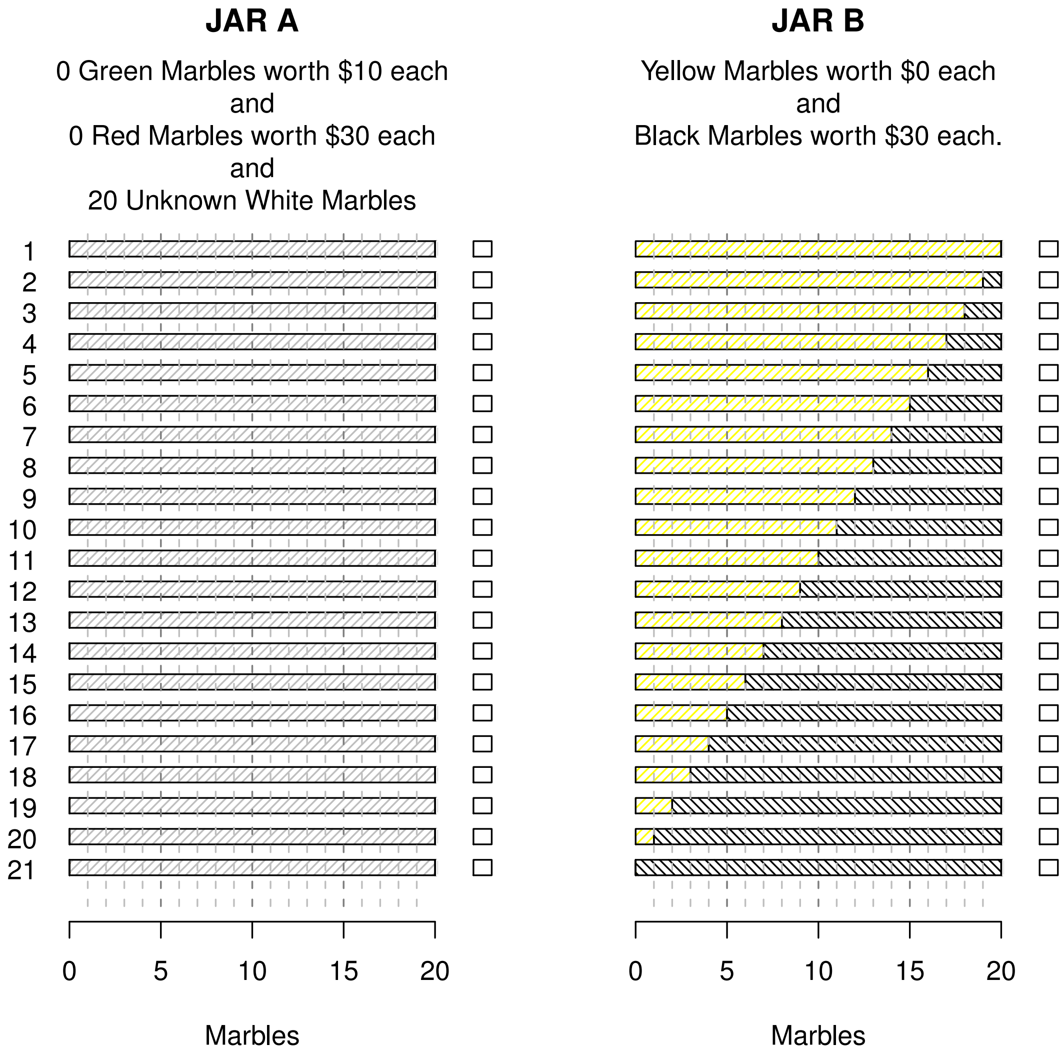
TASK 19

On this page you will make a series of decisions between two jars. JAR A will contain 0 Green marbles worth \$10 each, 5 Red marbles worth \$30 each, and 15 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



TASK 20

On this page you will make a series of decisions between two jars. JAR A will contain 0 Green marbles worth \$10 each, 0 Red marbles worth \$30 each, and 20 Unknown White marbles, which could be Red, Green or some mixture of the two. JAR B will contain Yellow marbles worth \$0 each and Black marbles worth \$30 each. For each row, decide whether you prefer to draw a marble from JAR A or from JAR B. Indicate your choice by checking the corresponding box. The contents of JAR A will remain the same in each row. JAR B will initially contain 20 Yellow \$0 Marbles and 0 Black \$30 marbles. The contents of JAR B will change. When you move down one row, one Yellow \$0 marble will be removed and replaced with one Black \$30 marble.



List of Figures

List of Tables

Ehrenwörtliche Erklärung

Das zweite Kapitel basiert auf einem gemeinsamen Projekt mit Georg Weizsäcker und Steffen Huck. Das dritte Kapitel basiert auf einem gemeinsamen Projekt mit James Andreoni und Charles Sprenger.

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertations benutzten Hilfmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Ort, Datum Unterschrift